

STIC-Biotech/ChemLib

From: Li, Ruixiang
Sent: Thursday, July 11, 2002 11:37 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/899,513

Please do a standard search on SEQ ID NOS: 1 and 2 against the commercial nucleic acid databases.

Thank you very much!

Ruixiang Li
GAU 1646
CM1 10E18
Mail Box-10C01
306-0282

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 7/12
Date Completed: 7/15
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 22:57:41 ; Search time 1813.39 Seconds

(Without alignments)
7131.725 Million cell updates/sec

Title: US-09-899-513-1

Perfect score: 618
Sequence: 1 atgggggctgccttgcgc.....ggtgctgcgcgcacacgiga 618

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 3595312

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

C	1	617.2	99.9	152947	2	AL354876	AL354876 Homo sapi
C	2	617.2	99.9	194215	2	AL607089	AL607089 Homo sapi
C	3	66	10.7	1116	6	AX068253	AX068253 Sequence
C	4	66	10.7	1119	6	AX076182	AX076182 Sequence
C	5	66	10.7	1119	6	AX138829	AX138829 Sequence
C	6	66	10.7	1119	9	HS272207	HS272207 Homo sapi
C	7	66	10.7	1119	9	AF411112	AF411112 Homo sapi
C	8	66	10.7	1325	6	AX168124	AX168124 Sequence
C	9	66	10.7	1560	6	AX352473	AX352473 Sequence
C	10	66	10.7	1720	6	AX068263	AX068263 Sequence
C	11	66	10.7	2444	6	AX305129	AX305129 Sequence
C	12	66	10.7	2480	6	AX128511	AX128511 Sequence
C	13	66	10.7	245880	2	AC079387	AC079387 Homo sapi
C	14	62	10.0	134940	2	AC018939	AC018939 Homo sapi
C	15	61.6	10.0	125020	2	AC183135	AC183135 Homo sapi
C	16	60	9.7	90005	2	AC106630	AC106630 Rattus no
C	17	58.4	9.4	1534	9	AF080214	AF080214 Homo sapi
C	18	58.4	9.4	4895	6	AR108643	AR108643 Sequence
C	19	58.4	9.4	4895	9	AF055917	AF055917 Homo sapi
C	20	58.4	9.4	4925	6	AX211769	AX211769 Sequence
C	21	58.4	9.4	11828	9	AF384819	AF384819 Homo sapi
C	22	58.2	9.4	248680	2	AC008737	AC008737 Homo sapi
C	23	58.2	9.4	78307	2	AC096500	AC096500 Rattus no
C	24	56.8	9.2	14662	1	SC663	SC663 Streptomy
C	25	56.4	9.1	39314	1	SGR300302	SGR300302 Streptomy
C	26	56.4	9.1	137007	2	AP004265	AP004265 Oryza sat
C	27	55.8	9.0	31226	1	SC6F7	SC6F7 Streptomy
C	28	55.8	9.0	155760	2	AC094490	AC094490 Rattus no
C	29	55.8	9.0	196662	2	AP001107	AP001107 Homo sapi
C	30	55.6	9.0	81673	2	AC016060	AC016060 Homo sapi
C	31	55.6	9.0	174546	2	AC087688	AC087688 Homo sapi
C	32	55.4	9.0	125020	2	AF429315	AF429315 Homo sapi
C	33	55	8.9	1842	5	AB042507	AB042507 Gallus ga
C	34	55	8.9	37445	1	SC8A6	SC8A6 Streptomy
C	35	54.8	8.9	127786	2	AC094129	AC094129 Rattus no
C	36	54.6	8.8	39896	1	SC2K31	SC2K31 Streptomy
C	37	54.6	8.8	132060	2	AC099243	AC099243 Rattus no
C	38	54.6	8.8	194177	2	AC021647	AC021647 Homo sapi
C	39	54.6	8.8	194855	2	AC093904	AC093904 Homo sapi
C	40	54.2	8.8	36816	2	AC109752	AC109752 Rattus no
C	41	54	8.7	39446	1	SC8F7	SC8F7 Streptomy
C	42	54	8.7	66597	2	AC090850	AC090850 Homo sapi
C	43	53.8	8.7	40235	1	SC210A7	SC210A7 Streptomy
C	44	53.4	8.6	85434	2	AC066610	AC066610 Homo sapi
C	45	53.4	8.6	186404	2	AC105307	AC105307 Bos tauru

ALIGNMENTS

RESULT 1
AL354876/c 152947 bp DNA linear HTG 27-JUN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-244H3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.

ACCESSION AL354876
VERSION AL354876.9 GI:14575158
KEYWORDS HTG; HTGS_Phrase1; HTGS_DRAFT.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 152947)

AUTHORS McLay, K.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Jun 28, 2001 this sequence version replaced gi:13334889.

Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: ba244h3
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 152070 bases at least Q40
 Consensus quality: 152318 bases at least Q30
 Consensus quality: 152484 bases at least Q20
 Insert size: 152747; sum-of-contrigs
 Insert size: 138665; 8.5% error; agarose-fp
 Quality coverage: 9.58x in Q20 bases; sum-of-contrigs Quality
 coverage: 10.56x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 54502: contrig of 54502 bp in length
 * 54503 54602: gap of 100 bp
 * 54603 99405: contrig of 44803 bp in length
 * 99406 99505: gap of 100 bp
 * 99506 152947: contrig of 53442 bp in length.

Location/Qualifiers
 1.152947
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-244H3"
 /clone.lib="RPCT-11.1"
 1.54502
 /note="assembly-fragment:02081
 fragment_chain:1"
 misc-feature
 54603..99405
 /note="assembly-fragment:01111
 fragment_chain:1"
 99506..152947
 /note="assembly-fragment:03694"
 misc-feature
 99506..152947

BASE COUNT 46288 a 33896 c 32934 g 33628 t 201 others
 ORIGIN

Query Match 99.9%; Score 617.2; DB 2; Length 152947;
 Best Local Similarity 99.8%; Pred. No. 2.9e-82;
 Mismatches 617; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 atgaggagctccttctgtcgtacgctcgcagtaacctctctctgcacattcaagttca 60
 |||||
 Db 54085 ATGGGGGCTCCCTTTGCGTAGCGTCCGACGTAACCTTTCTTGCCACTTCAAGTCA 54026
 |||||
 QY 61 gagatgaacagcagtgctgtggagcagtggtgtgtggcgagctgcacgtctggatgacct 120
 |||||
 Db 54025 GAGATGAACAGCAGTGTGTGGGACCTGGGTGTGGCGGCTGCAAGCCCTCTGGATGACCTT 53966
 |||||
 QY 121 gcttcgtctctgtgtggcgagcagctatgctgtgacctggagcctggggcttgcagcc 180
 |||||
 Db 53965 GCTTCGCTTCAATCGTGGTGGCCGCGGCTATGCTTGGCACTGGGCGCTGGGCGTGCACGCC 53906
 |||||
 QY 181 aacgtggcgagcctgtgaaatgtcatcgcagcgagcgagcctggcgagccttgcct 240
 |||||
 Db 53905 AACGTGGCGGCGCTGCGCAATGTTATCCGACGCGGCGGCGCTGGGCGCAAGCCCTTCTT 53846
 |||||
 QY 241 ctctaccgtgtcaacctgtgctgtgtatgtagtcttctcaagctacagctgtcagctgtg 300
 |||||
 Db 53845 CTCTACCTGTTCAACCTGTGCTGTGATGAGTCTTCAACGCTCACGCTGCACACTGTGG 53786
 |||||
 QY 301 ctcaaccctacacgtggcgagcggcgagcgagcgtgtgcacgagcgagcgagcgagcact 360
 |||||

Db 53785 CTCACCTACTACTGGGCGCTGGCCGAGGCGCCCTCTGCCACGCGCGCGGCGCACTACT 53726
 |||||
 QY 361 acgtgtcacactatgtcgagtggtgtcttcgcgcgctcatcaagctgtgctgagct 420
 |||||
 Db 53725 ACCTGTCCACCTATGCGGCGGCTGTCTTCCGCGGCTCATCAGTGTGCGCTGCGGCT 53666
 |||||
 QY 421 tcttaagcgggtcccgagcgagcgagctgtgcccgtctgtgagcgagtgagcgagcg 480
 |||||
 Db 53665 TCTTACGCGGCTCCCGGCGGCGGCGGCTGCGGCTGCGGCTGCTTACGCGGCGGCG 53606
 |||||
 QY 481 cgcgcgtctgtcgcctctgtcgtctgtgagcgagcctgtgagcctgtgagcagc 540
 |||||
 Db 53605 CCGCGCTGCTGCTGCGCTTGTGCTGTGCTGCGGCGCTGCGGCTGCTGCTGAGAGAC 53546
 |||||
 QY 541 gctgggcaagctcgagtgctgctcgcgcagctgtgagctgtgagcctgtgagcctgtg 600
 |||||
 Db 53545 GCTGGGCAACCTGCGGCGGCTGCTGCGGCGGCTGCTGCGGCGGCTGCTGCTGCTG 53486
 |||||
 QY 601 tcttcgagcgagcagctgtga 618
 |||||
 Db 53485 TGCTCGCGGCGCAACGTGA 53468
 |||||

RESULT 2
 AL607089/c 194215 bp DNA linear HTG 30-JAN-2002
 LOCUS Homo sapiens chromosome 1 clone RP11-24819, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 2 ordered pieces.
 ACCESSION AL607089
 VERSION AL607089.15 GI:18476794
 KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (sites)
 REFERENCE Brown, J.
 AUTHORS Direct Submission
 TITLE Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 JOURNAL Cambridgehire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Feb 1, 2002 this sequence version replaced gi:17402315.
 COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: ba244h3
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 192819 bases at least Q40
 Consensus quality: 193115 bases at least Q30
 Consensus quality: 193255 bases at least Q20
 Insert size: 194215; sum-of-contrigs
 Insert size: 193590; 7.2% error; agarose-fp
 Quality coverage: 9.54x in Q20 bases; sum-of-contrigs Quality
 coverage: 9.62x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contrigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 16618: contrig of 16618 bp in length
 * 16619 16718: gap of 100 bp
 * 16719 194215: contrig of 177497 bp in length.

LOCUS	AX168124	1325 bp	DNA	Linear	PAT 03-JUL-2001
DEFINITION	Sequence 42 from Patent W00142288.				
ACCESSION	AX168124				
VERSION	AX168124..1	GI:14597402			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 1325)				
AUTHORS	Buford,N., Baughn,M.R., Au-Young,J., Yang,J., Lu,D.A. and Reddy,R.				
TITLE	G-protein coupled receptors				
JOURNAL	Patent: WO 0142288-A 42 14-JUN-2001;				
	Incyte Genomics, Inc. (US)				
FEATURES	Location/Qualifiers				
Source	1..1325				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/note="Incyte ID No: 3168839CB1"				
BASE COUNT	195 a 473 c 403 g 254 t				
GCIN					
Query Match	10.7%; Score 66; DB 6; Length 1325;				
Best Local Similarity	50.2%; Pred. No. 0.72;				
Matches 223; Conservative	1; Mismatches 207; Indels 13; Gaps 2;				
Oy	112 gatgacctgtcgcgttcatggtgcccgcgcctatgccttgacatgcttgagc 171				
Db	90 GACTACCGACCTACCCACGCGCTGCACATTGGGTGCTACAGCTTGCTGGCCGG 149				
Oy	172 ctgcagcaacagctggcgcctggcgaaltgtcatlccgcagcgcgcgcctggccag 231				
Db	150 CTCGCCCTCAACGCGCTAGCCCT--CTGGGCTCTTCGCGCGCGCTCGGCACATCG 206				
Oy	232 gccgcctcttactctttaaactggtcctcgtgtgtagttcttaagctaagctc 291				
Db	207 GGGGTACGCTACATGTGTAACCTGGCGGCCACGACGACTCTCTTACCCCTTCGCTG 266				
Oy	292 cagctgtgtctaacctactacgt-----ggcctggccggagagccgcctgcac 341				
Db	267 CCCGTCGCTCTCTACTACGCACTGGACCACTGGCCCTTCCCGACCTCCTGGCCAG 326				
Oy	342 gcgcgcggggccaactactaagctgtccacactatgcgcgcggtgtcttcgcgcgtcatc 401				
Db	337 ACGAGGGCGCGCATTTCCAGATGAACATGTAGCGCACACCTGCATCTCTGATGCTCAAC 386				
Oy	402 agcgtgtgcgcgtggcgtgtgtaacgcggtcccgccagggccgcgcgccttgcc 461				
Db	387 AACGTGACCGCTACGCCGCCCATGTGTGACACCGCTGCGACTGCGCACCTGGCGGGCCC 446				
Oy	462 cgtgtcctaagcgbccgcgcgcgcgcgtgccttgcccttgcttgagcgttgagcgttgcc 521				
Db	447 CGCGTGGCGCGCTCTCTGCTGGCGGTGTGGCGCATCTGTGTTTCCGTCGCC 506				
Oy	522 cctccttgcttgagcacgcgtcg 545				
Db	507 GCCGCCCGGTGACAGGCCCTCG 530				
RESULT 9	AX352473	1560 bp	DNA	Linear	PAT 06-FEB-2002
LOCUS	AX352473				
DEFINITION	Sequence 25 from Patent W00190187.				
ACCESSION	AX352473				
VERSION	AX352473..1	GI:18617722			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (sites)				

AUTHORS
Pedigaru, M., Spytek, K.A., Majumder, R., Tchiernev, V.T., Grosse, N.M.,
Szerkner, E.D., Alshook, J.P., Burgess, C.E., Shmukets, R.A.,
Tapiher, R.J., Casman, S.J., Gangoli, E., Macdougall, J.R., Stone, D.J.
and Smithson, G.

TITLE
Novel proteins and nucleic acids encoding same
Patent: WO 0190187-A-25 29-NOV-2001;

JOURNAL
Curagen Corporation (US)

FEATURES
source Location/Qualifiers
1..1560
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT
230 a 561 c 447 g 322 t

ORIGIN

```

Query Match      10.7%; Score 66; DB 6; Length 1560;
Best Local Similarity 50.2%; Pred. No. 0.7;
Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

OY   112 gatgaacctgtctgcattcatcgttgtagcccgagcactatgcttggcaactggagctggagg 171
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    357 GACTACGAGCACTAACCCAGCGCCTGCATGTGGTGCTTAACACTTGTGTCTGGTCGCCGGG 416

OY   172 ctgccagccaacagtggcgcccttggcaatgttcatacgcagcgcgagcgagcgtctggccag 231
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    417 CTCGCCCTCAACGGCGCTAAGCCT--CTGGGTCTTCTCGCGCGCGCTGGCGGTGCATCG 473

OY   232 gccctgtctctacactgtttacaactgcttgtgtgtatgatgtcttcttaacgtcaagctg 291
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    474 GTGGTAGACCGGTACATTGTAACTGGCGGGCCAGACACGTGCTTTACCCCTCTCGCTG 533

OY   292 cagctgtgctaacactactactcy-----gacctggccggaggccgacctggcac 341
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    534 CCCGTTCTCTCTCTCTCTACTACGACACTGCACCACTGGGCCCTTCCCAGACCTCTGTGCGAG 593

OY   342 gcggcgccggagccaccactaatcgtttccacctatgcgagcggtgtcttgcgcgcatac 401
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    594 ACGACGGCGCCCATTTCCAGATGAACAATGTACGGCACCTGCATCTTCTCATGTCTATC 653

OY   402 agcgtgtgcgcgtcgagcttcgtacgcggtlcccgaggccaaaggcgagctgcgcgcctggcc 461
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    654 AACGTGAGACCGCTACGGCGCATGCTGTGACCCGCGTGCAGTGCACCACCTGCAGCGGCC 713

OY   462 cggtgtccaaagbbcccccgcgcggtctcgcgcgcttcgcgtcttgctggcgagctggcc 521
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    714 CGCGTGGCGCGGCTGCTTGCCTGGGCGCTGTGGGCGCTCATCTGTGTGTCGCGTCC 773

OY   522 cctcccctgcctggagcaacgctgg 545
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    774 GCCGCCCGCGTGCACAGGCCCTCG 797

RESULT 10
AX068263 1720 bp DNA linear PAT 25-JAN-2001
LOCUS AX068263 Sequence 25 from Patent W00102563.
DEFINITION AX068263
ACCESSION AX068263
VERSION AX068263.1 GI:12578460
KEYWORDS
SOURCE
ORGANISM human.
human.sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1720)
AUTHORS Kato,S. and Kimura,T.
TITLE Human proteins having hydrophobic domains and dnas encoding these
proteins
PATENT: WO 0102563-A-25 11-JAN-2001;
SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc.. (JP)
JOURNAL Location/Qualifiers
FEATURES
source 1..1720
/organism="Homo sapiens"
/db_xref="taxon:9606"
```


*	207064	209081:	contig of 2036 bp in length
*	209082	209181:	gap of unknown length
*	209182	210296:	contig of 1115 bp in length
*	210297	210396:	gap of unknown length
*	210397	212804:	contig of 2408 bp in length
*	212805	212904:	gap of unknown length
*	212905	214017:	contig of 1113 bp in length
*	214018	214117:	gap of unknown length
*	214118	216475:	contig of 2358 bp in length
*	216476	216575:	gap of unknown length
*	216576	217968:	contig of 1333 bp in length
*	217969	218068:	gap of unknown length
*	218069	219095:	contig of 1027 bp in length
*	219096	219195:	gap of unknown length
*	219196	220470:	contig of 1275 bp in length
*	220471	220570:	gap of unknown length
*	220571	221822:	contig of 1252 bp in length
*	221823	221922:	gap of unknown length
*	221923	223428:	contig of 1506 bp in length
*	223429	223528:	gap of unknown length
*	223529	224826:	contig of 1228 bp in length
*	224827	224926:	gap of unknown length
*	224927	226124:	contig of 1198 bp in length
*	226125	226224:	gap of unknown length
*	226225	227922:	contig of 1698 bp in length
*	227923	228022:	gap of unknown length
*	228023	229064:	contig of 1042 bp in length
*	229065	229164:	gap of unknown length
*	229165	230519:	contig of 1355 bp in length
*	230520	230649:	gap of unknown length
*	230650	231847:	contig of 1228 bp in length
*	231848	231947:	gap of unknown length
*	231948	232513:	contig of 1566 bp in length
*	232514	232613:	gap of unknown length
*	232614	234860:	contig of 1247 bp in length
*	234861	234960:	gap of unknown length
*	234961	235968:	contig of 1008 bp in length
*	235969	236068:	gap of unknown length
*	236069	237121:	contig of 1053 bp in length
*	237122	237221:	gap of unknown length
*	237222	238257:	contig of 1036 bp in length
*	238258	238357:	gap of unknown length
*	238358	239413:	contig of 1056 bp in length
*	239414	239513:	gap of unknown length
*	239514	240850:	contig of 1337 bp in length
*	240851	240950:	gap of unknown length
*	240951	242140:	contig of 1190 bp in length
*	242141	242240:	gap of unknown length
*	242241	243504:	contig of 1264 bp in length
*	243505	243604:	gap of unknown length
*	243605	244749:	contig of 1145 bp in length
*	244750	244849:	gap of unknown length
*	244850	245880:	contig of 1031 bp in length

BASE COUNT	61016	a	60422	c	59511	g	60793	t	4138	others
ORIGIN										

Query Match 10.7%; Score 66; DB 2; Length 245880;
 Best Local Similarity 50.2%; Pred. NO. 0.25;
 Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

[illegible]

Db	81614	CTCCCCCTCAAGGCGCTAAGCCCT---CTGGGGTCTTCAGCGGGCGCTCGGGGTGCATCTGG	81558
OY	232	gccctgctctctctaccgtgttcaacctggtcgtgtgtatgtatgtcttcaagctcaagctg	291
Db	81557	GTGTGAGACCGTGTATCATCTGTAACTGTGGCGGCGACAGCACTGCCTCTTCACCCCTGCGTGG	81498
OY	252	cagctggtgctctaccattactcgt-----ggctctggcccggaagcgccgtcgccac	341
Db	81497	CCCGTTCCTCTCTCTCTACTATAGCGACATGCAACACTGCGCCCTTCCCGGACCTCTGTGGCCAG	81438
OY	342	ggcgccgggagcccaactactacgtgtgcacactatgctggcggtgtctctgcgcgctcaac	401
Db	81437	ACGACGGGCGCCATTTTCCAGATGAACATGTAAAGCGACACTGCATTTTCTGTATGCTCATTC	81378
OY	402	agcgtgtccgctgctggcgtcttgtaacgcggtctccggagcccaaggcgagctgcgcgcctgagcc	461
Db	81377	AACGTGAGCAACCGTAAAGCGGCCATCTGTGCAACCGCGATGGCACTGCGCCACCTGCGGGGGGCC	81318
OY	462	cgggtgctcaagbgbccccgcgcgcgcgtgcgtgccttcgccttgctgctgagcgagctgtgac	521
Db	81317	CGCGTGGCGCGGCTCTCTGCGCTGGGCGGTGGGGCGCTCATCTGTGTTTGCCGTGCCCC	81258
OY	522	cctccctccttgtagaacacgcgtcg	545
Db	81257	GCCTCCGCGGTGACAGGCGCTCTGG	81234

RESULT	14
AC018939	
LOCUS	AC018939 134940 bp DNA 1 linear HTG 13-JUL-2000
DEFINITION	Homo sapiens clone RP11-21D12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC018939
VERSION	AC018939..3 GI:9123989
KEYWORDS	HTG; HTGS_PHASED.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 134940)
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens, clone RP11-21D12
REFERENCE	unpublished
AUTHORS	2 (bases 1 to 134940)
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

TITLE	Direct Submission
JOURNAL	Submitted (24-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Jul 13, 2000 this sequence version replaced gi:6970649.

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information

Center project name: U3982
Center clone name: 21_D_12

***** NOTE: This record contains 141 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 877: contig of 877 bp in length
* 878 977: gap of 100 bp
* 978 1861: contig of 884 bp in length
* 1862 1961: gap of 100 bp
* 1962 2860: contig of 899 bp in length
* 2861 2960: gap of 100 bp
* 2961 3820: contig of 860 bp in length
* 3821 3920: gap of 100 bp
* 3921 4784: contig of 864 bp in length
* 4785 4884: gap of 100 bp
* 4885 5793: contig of 909 bp in length
* 5794 5893: gap of 100 bp
* 5894 6780: contig of 887 bp in length
* 6781 6880: gap of 100 bp
* 6881 7754: contig of 874 bp in length
* 7755 7854: gap of 100 bp
* 7855 8732: contig of 878 bp in length
* 8733 8832: gap of 100 bp
* 8833 9710: contig of 878 bp in length
* 9711 9810: gap of 100 bp
* 9811 10670: contig of 860 bp in length
* 10671 10770: gap of 100 bp
* 10771 11658: contig of 888 bp in length
* 11659 11758: gap of 100 bp
* 11759 12653: contig of 895 bp in length
* 12654 12753: gap of 100 bp
* 12754 13647: contig of 894 bp in length
* 13648 13747: gap of 100 bp
* 13748 14629: contig of 882 bp in length
* 14630 14729: gap of 100 bp
* 14730 15598: contig of 869 bp in length
* 15599 15698: gap of 100 bp
* 15699 16583: contig of 885 bp in length
* 16584 16683: gap of 100 bp
* 16684 17579: contig of 896 bp in length
* 17580 17679: gap of 100 bp
* 17680 18563: contig of 884 bp in length
* 18564 18663: gap of 100 bp
* 18664 19542: contig of 879 bp in length
* 19543 19642: gap of 100 bp
* 19643 20506: contig of 864 bp in length
* 20507 20606: gap of 100 bp
* 20607 21487: contig of 881 bp in length
* 21488 21587: gap of 100 bp
* 21588 22482: contig of 895 bp in length
* 22483 22582: gap of 100 bp
* 22583 23475: contig of 893 bp in length
* 23476 23575: gap of 100 bp
* 23576 24475: contig of 900 bp in length
* 24476 24575: gap of 100 bp
* 24576 25473: contig of 898 bp in length
* 25474 25573: gap of 100 bp
* 25574 26472: contig of 899 bp in length
* 26473 26572: gap of 100 bp
* 26573 27450: contig of 878 bp in length
* 27451 27550: gap of 100 bp
* 27551 28440: contig of 890 bp in length
* 28441 28540: gap of 100 bp
* 28541 29431: contig of 891 bp in length

29432 29531: gap of 100 bp
* 29532 30392: contig of 861 bp in length
* 30393 30492: gap of 100 bp
* 30493 31363: contig of 871 bp in length
* 31364 31463: gap of 100 bp
* 31464 32327: contig of 864 bp in length
* 32328 32427: gap of 100 bp
* 32428 33268: contig of 841 bp in length
* 33269 33368: gap of 100 bp
* 33369 34264: contig of 896 bp in length
* 34265 34364: gap of 100 bp
* 34365 35235: contig of 871 bp in length
* 35236 35335: gap of 100 bp
* 35336 36212: contig of 877 bp in length
* 36213 36312: gap of 100 bp
* 36313 37171: contig of 859 bp in length
* 37172 37271: gap of 100 bp
* 37272 38141: contig of 870 bp in length
* 38142 38241: gap of 100 bp
* 38242 39098: contig of 857 bp in length
* 39099 39198: gap of 100 bp
* 39199 40093: contig of 895 bp in length
* 40094 40193: gap of 100 bp
* 40194 41082: contig of 889 bp in length
* 41083 41182: gap of 100 bp
* 41183 42041: contig of 859 bp in length
* 42042 42141: gap of 100 bp
* 42142 43033: contig of 892 bp in length
* 43034 43133: gap of 100 bp
* 43134 43975: contig of 842 bp in length
* 43976 44075: gap of 100 bp
* 44076 45026: contig of 951 bp in length
* 45027 45126: gap of 100 bp
* 45127 46058: contig of 932 bp in length
* 46059 46158: gap of 100 bp
* 46159 47015: contig of 857 bp in length
* 47016 47115: gap of 100 bp
* 47116 48006: contig of 891 bp in length
* 48007 48106: gap of 100 bp
* 48307 48978: contig of 872 bp in length
* 48979 49078: gap of 100 bp
* 49079 49987: contig of 909 bp in length
* 49988 50087: gap of 100 bp
* 50088 50999: contig of 912 bp in length
* 51000 51099: gap of 100 bp
* 51100 51988: contig of 889 bp in length
* 51989 52088: gap of 100 bp
* 52089 52964: contig of 876 bp in length
* 52965 53064: gap of 100 bp
* 53065 53967: contig of 903 bp in length
* 53968 54067: gap of 100 bp
* 54068 54953: contig of 886 bp in length
* 54954 55053: gap of 100 bp
* 55054 55944: contig of 891 bp in length
* 55945 56044: gap of 100 bp
* 56045 56921: contig of 877 bp in length
* 56922 57021: gap of 100 bp
* 57022 57901: contig of 880 bp in length
* 57902 58001: gap of 100 bp
* 58002 58880: contig of 879 bp in length
* 58881 58980: gap of 100 bp
* 58981 59882: contig of 902 bp in length
* 59883 59982: gap of 100 bp
* 59983 60870: contig of 888 bp in length
* 60871 60970: gap of 100 bp
* 60971 61852: contig of 882 bp in length
* 61853 61952: gap of 100 bp
* 61953 62814: contig of 862 bp in length
* 62815 62914: gap of 100 bp
* 62915 63746: contig of 832 bp in length
* 63747 63846: gap of 100 bp
* 63847 64699: contig of 853 bp in length
* 64700 64799: gap of 100 bp

Mon Jul 15 14:32:23 2002

us-09-899-513-1.rge

Page 12

Search completed: July 13, 2002, 02:12:56
Job time: 11715 sec

PI Glucksmann MA, Welch NS;
 XX WPI: 2000-339687/29.
 DR P-PSDB: AAV79564.

PT G-protein coupled receptors and the nucleic acids that encode them
 XX useful for treating anemia, neutropenia and/or thrombocytopenia -
 XX
 PS Claim 3: Page 138-139; 140pp; English.

CC The present sequence is that of cDNA coding for the novel human
 CC G protein coupled receptor (GPCR) 15334, a protein that participates
 CC in signalling pathways and which is expressed in colon, pancreas,
 CC tonsil, lymph node, spleen, thymus, adrenal gland, heart and
 CC peripheral blood lymphocytes, including megakaryocytes and
 CC erythroblasts. To obtain 15334 cDNA, an expressed sequence
 CC tag (EST) was selected on the basis of homology to GPCR sequences.
 CC Primers based on the EST were used to identify 15334 cDNA from a
 CC spleen cDNA library. Positive clones were sequenced and the
 CC overlapping fragments assembled. The invention relates to the
 CC discovery of novel GPCRs, and the polynucleotides encoding them,
 CC as well as to methods of using the GPCR polynucleotides and
 CC polynucleotides as targets for diagnosis and treatment of
 CC receptor-mediated disorders. It also relates to drug screening
 CC methods using the polypeptides and polynucleotides to identify
 CC agonists and antagonists for diagnosis and treatment. Methods
 CC of treating a disorder by modulating the level or activity of a
 CC GPCR nucleic acid, including 15334 nucleic acid, are also claimed,
 CC where the disorder is selected from anaemia, neutropenia and
 CC thrombocytopenia (claimed). Other conditions associated with
 CC inappropriate GPCR expression and activity can also be treated,
 CC e.g. retinitis pigmentosa, nephrogenic diabetes insipidus and
 CC other disorders of the central and peripheral nervous systems.

CC Sequence 2559 BP; 506 A; 785 C; 648 G; 620 T; 0 other;

Query Match 10.9%; Score 67.6; DB 21; Length 2559;
 Best Local Similarity 50.5%; Pred. No. 0.00023;
 Matches 224; Conservative 1; Mismatches 206; Indels 13; Gaps 2;

QY 112 gatgacctctgctgcttcacgtgtgctccgcgacatacgtgctgagcgctggg 171
 DB 185 gactacgacactaccacccgcgctgactgtgtgtctacagctgtgtgctgctggc 244
 QY 172 ctgcagcagcaactgtgctgctgcaatgttcaatccgagcgcgctggcgccag 231
 DB 245 ctccctcctcaagcgctagccct--ctgggtcttccctgcgcgctgcgctgacac 301
 QY 232 gacctgtctctacgttcaactgctgtgtgtatgtctcttcacgtcagctg 291
 DB 302 gtgtgtgagcgtgtacatgtgttaactgtgcgagcagactgctcttaccctctgctg 361
 QY 292 cagctgtgacactacacttactg-----ggcctgagccgagcgagcgctgacac 341
 DB 362 ccgcttctctctctactactaagcactgacacactggcctctcccgactctctgtgacag 421
 QY 342 gcggcgaggggacactactacgtgtccaactatgcgagtgatcttccgcgcgtatc 401
 DB 422 acgagcgagcgacatctcagatgaacatgtacgagcagctgcatctctcgtatgcatc 481
 QY 402 agcgtgtgacgtcgtgcttctgtaacggtgtcccgagcccaaggcgagcgctggcc 461
 DB 482 aacgtgagcagcgtacgcgagcgtgacacccgctgacgtgcacactgtcgagggcgc 541
 QY 462 cgggtgacactagggggcgccgcgagcgctgctgagccttgcgtgctgagcgagctggcc 521
 DB 542 cggctgtgagcgagcgtctgctgctggcgctgtgagcgctacatccgtgtgttgcgctgccc 601
 QY 522 cctccctgctgagcgacacgtg 545
 DB 602 gcgcgcgcgctgacagggcctcg 625

RESULT 2

AAF28683
 ID- AAF28683 standard; cDNA; 1116 BP.

AC AAF28683;

DT 05-APR-2001 (first entry)

DE Human protein HP03378 coding sequence #1.

KW Human; hydrophobic domain; immune deficiency; autoimmune disorder;
 KW allergy; tissue growth; regeneration; wound healing; burn; tumour;
 KW periodontal disease; thrombolytic condition; haemostatic condition;
 KW infection; ss.

OS Homo sapiens.

PN MO200102563-A2.

PD 11-JAN-2001.

PF 16-JUN-2000; 2000MO-UP03943.

PR 02-JUL-1999; 99JP-0188835.

PA (SAGA) SAGAMI CHEM. RES. CENT.
 (PROT-) PROTEGENE INC.

PI Kato S, Kimura T;

DR WPI: 2001-071581/08.

DR P-PSDB: AAB61612.

PT New human proteins with hydrophobic domains, useful for the treatment
 PT of immune disorders, tumors, allergic conditions, thrombosis and
 PT microbial infection.

PS Claim 3; Pages 116-117; 153pp; English.

CC The present invention relates to human proteins (AAB61608-AAB61617) and
 CC their coding sequences (AAF28679-AAF28698). The proteins of the present
 CC invention have hydrophobic domains and can be used for the treatment of
 CC various immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, graft-versus-host disease and
 CC Guillain-Barre syndrome. The proteins may also be useful in the treatment
 CC of allergic reactions and conditions, such as asthma and in regulation of
 CC haematopoiesis or lymphoid cell deficiencies. The proteins may also have
 CC utility in compositions used for bone, cartilage, tendon and/or nerve
 CC tissue growth or regeneration as well as wound healing and in the
 CC treatment of burns. The proteins may be used in the treatment of
 CC periodontal disease and in other tooth repair processes. Other uses
 CC include treatment of thrombolytic and haemostatic conditions, treatment
 CC of prevention of tumours and inhibiting infection by bacteria, viruses,
 CC fungi and other parasites.

SO Sequence 1116 BP; 141 A; 420 C; 339 G; 216 T; 0 other;

Query Match 10.7%; Score 66; DB 22; Length 1116;
 Best Local Similarity 50.2%; Pred. No. 0.00046;
 Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

QY 112 gatgacctctgctgcttcacgtgtgctccgcgacatacgtgctgagcgctggg 171
 DB 49 gactacgacactaccacccgcgctgactgtgtgtctacagctgtgtgctgctggcgg 108
 QY 172 ctgcagcagcaactgtgctgctgcaatgttcaatccgagcgcgctggcgccag 231
 DB 109 ctccctcctcaagcgctagccct--ctgggtcttccctgcgagcgctgctggtgacacg 165
 QY 232 gacctgtctctactactgttcaactgctgctgtgtgtatgttcttcaacgtcagctg 291

Db 166 gtgtgtgagcgtgtacatgtgttaacctgtgagcgagcagcagccttccacctctgtgt 225
 QY 292 cagctgtgtgtacactactactactcgtg-----gacctgcccggagggcgacctccac 341
 Db 226 ccgcttctgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 285
 QY 342 gggcgccggggccacactactactactactactactactactactactactactactactact 401
 Db 286 acgacggggggcgcacatcttccagatgaacatgaacagcagcagcagcagcagcagcagc 345
 QY 402 agcgtgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgt 461
 Db 346 aacgtgtgacccgtacgcccgcacatcgtgtaccccgctgtgacgtgtgacgtgtgacgtgt 405
 QY 462 cgtgtgtacaggggccccggcgcgcgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 521
 Db 406 cgcgtgtgagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 465
 QY 522 cctccctgctgtgagcagcgtgtg 545
 Db 466 gccgcccgcgtgtcacagcgccctgtg 489

RESULT 3

ABA02176
 ID ABA02176 standard; cDNA; 1116 BP.

AC ABA02176;

DT 08-FEB-2002 (first entry)

DE Human G protein-coupled receptor TGR4-encoding cDNA.

KW Human; TGR4; G protein-coupled receptor; GPCR; 7TM receptor; brain;
 KW signal transduction; ligand screening; drug discovery; diagnosis;
 KW central nervous system; circulatory system; digestive system;
 KW immunological disease; inflammation; cancer; diabetes; cerebroprotective;
 KW cardiant; immunomodulator; cytostatic; antiinflammatory; antidiabetic;
 KW ss.

OS Homo sapiens.

FH Key Location/Qualifiers
 FT 1.1116
 FT CDS /tag= a
 FT /partial
 FT /product= "Human TGR4"
 FT /note= "No stop codon given in the specification"

PN WO200177326-A1.

PD 18-OCT-2001.

PF 12-APR-2001; 2001WO-JP03144.

PR 12-APR-2000; 2000JP-0110761.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Hikiichi Y, Matsui H, Shintani Y;

DR WPI: 2002-025901/03.

DR P-PSDB; AAM52650.

PT Human brain-originated G protein-coupled receptor protein TGR4,
 PT applicable in diagnosis and developing drugs for diseases of e.g.
 PT central nervous system, digestive organs, immunological diseases,
 PT inflammation, cancer and diabetes

PS Claim 5; Page 92; 99pp; Japanese.

CC This sequence represents cDNA encoding a novel human brain-derived G

CC protein-coupled receptor (GPCR), designated TGR4. Like all GPCRs, TGR4
 CC has 7 putative transmembrane domains and is likely to be involved in
 CC signal transduction. The invention also relates to vectors and host cells
 CC comprising TGR4 nucleotides; the recombinant expression of TGR4; an
 CC antibody against TGR4; methods for quantitating TGR4 protein or nucleic
 CC acids; and methods of screening for TGR4 ligands or modulators of TGR4
 CC activity or expression. The TGR4 protein, nucleotides encoding it, and
 CC antibodies against TGR4 may be used in the diagnosis of diseases of the
 CC central nervous system, circulatory system and digestive system;
 CC immunological diseases; inflammation; cancer; and diabetes. They may also
 CC be used to develop drugs for the treatment of the above conditions.
 CC XX

Sequence 1116 BP; 141 A; 420 C; 339 G; 216 T; 0 other;

Query Match 10.78; Score 66; DB 24; Length 1116;
 Best Local Similarity 50.28; Pred. No. 0.00046;
 Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

QY 112 gatgaccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 171
 Db 49 gactaccgacctaccaccacgacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 108
 QY 172 ctggcagccacagctgtgcccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 231
 Db 109 ctccccctcaacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 165
 QY 232 gccctgtctcttactactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 291
 Db 166 gtgtgtgagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 225
 QY 292 cagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 341
 Db 226 ccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 285
 QY 342 gggcgccggggccacactactactactactactactactactactactactactactactactact 401
 Db 286 acgacggggggcgcacatcttccagatgaacatgaacagcagcagcagcagcagcagcagcagc 345
 QY 402 agcgtgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgcccgtgccc 461
 Db 346 aacgtgtgacccgtacgcccgcacatcgtgtaccccgctgtgacgtgtgacgtgtgacgtgtgac 405
 QY 462 cgtgtgtacaggggccccggcgcgcgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 521
 Db 406 cgcgtgtgagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 465
 QY 522 cctccctgctgtgagcagcgtgtg 545
 Db 466 gccgcccgcgtgtcacagcgccctgtg 489

RESULT 4

AAA64367
 ID AAA64367 standard; DNA; 1119 BP.

AC AAA64367;

DT 20-DEC-2000 (first entry)

DE DNA encoding a P2Y-like 7 transmembrane receptor designated AXOR17.

KW Human; P2Y-like protein; 7 transmembrane receptor; AXOR17; infection;
 KW pain; cancer; diabetes; obesity; anorexia; bulimia; asthma; stroke;
 KW Parkinson's disease; acute heart failure; hypertension; hypotension;
 KW urinary retention; osteoporosis; angina pectoris; myocardial infarction;
 KW ulcer; benign prostatic hypertrophy; migraine; vomiting; schizophrenia;
 KW psychotic disorder; neurological disorder; depression; dyskinesia;
 KW Huntington's disease; Gilles de la Tourette's syndrome; vaccine; ss.

PS Homo sapiens.

CC Key Location/Qualifiers

[illegible]

Db	226	ccgcctctgtctctctctactacgacatgcacacatgtgcctctcccgacactctcgtgtccag	285
Qy	342	gcgcgcggggccacactactactacgtgtccacactatgcgcggttggtctctgcgcgcgtcacc	401
Db	286	acgacgcggcgccactctccacatgacatgtacgcagcgtcatctctcccgatgtctacc	345
Qy	402	agcgtgtgcgcgtgcgcgtctctgtacgcggtctccgggcccagggcgcgtgtgcgcgtctgcc	461
Db	346	aacgtgtgaccgcgtacgcgcgcacatgtgacacccgcctgcgactgcgcacactgcgcgcgcc	405
Qy	462	cgctgcgtacagbpbcccccgcgcgcgcgtgcgtgcctctgcctctgcgtgcgcgcgtgcgc	521
Db	406	cgcgtgtgcgcgcgtctgtctctgcctgcgcgcgtgtgtggcgctcatctctgtgttgcgcgtgcc	465
Qy	522	cctccctgccttgagacacgcgttg	545
Db	466	gcgcgcgcgcgtgacagcgcctcg	489
RESULT 9			
AA	98045		
XX	AA	98045 standard; DNA; 1119 BP.	
XX	AA	98045;	
AC			
XX	12-MAR-2002	(first entry)	
DE			
XX		Human DNA for potential G protein-coupled receptor #3.	
KW		Human; G protein-coupled receptor; GPCR; ds; GAT4; galanin receptor;	
KW		Alzheimer's disease; amyotrophic lateral sclerosis; asthma;	
KW		atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;	
KW		chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;	
KW		depression; epilepsy; macular degeneration; lymphoma; melanoma;	
KW		multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;	
KW		psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;	
KW		tuberculosis; cognition disorder; memory disorder; anorexia;	
KW		hormonal release disorder; cardiovascular activity disorder;	
KW		pain perception disorder; obesity; diabetes; obesity;	
KW		diabetes; hyperlipidaemia; stroke; gene therapy.	
XX			
OS		Homo sapiens.	
PN		WO200185791-A1.	
XX			
PD		15-NOV-2001.	
XX			
PF		11-MAY-2001; 2001WO-US15332.	
XX			
XX		11-MAY-2000; 2000US-203217P.	
PR		18-MAY-2000; 2000US-205945P.	
XX			
PA		(LIFE-) LIFESPAN BIOSCIENCES INC.	
XX			
PI		Brown JP, Miller M, Burner G, Fabre-Suver C, Pritchard D;	
XX			
DR		WPI; 2002-066595/09.	
XX			
PT		Novel G protein-coupled receptor polypeptides including galanin	
PT		receptor polypeptides useful for identifying modulators that are useful	
PT		for treating Alzheimer's disease, psoriasis, melanoma, multiple	
PT		sclerosis, stroke	
XX			
PS		Disclosure; Page 113-114; 144pp; English.	
XX			
CC		The invention relates to an isolated polypeptide encoded by a	
CC		nucleic acid molecule that is at least 80% identical to the G	
CC		protein-coupled (GPCR) polynucleotides included in the specification.	
CC		Also included are probes based on the GPCR sequences (including	
CC		antisense probes), a host cell comprising an expression vector comprising	
CC		the GPCR sequence, antibodies raised against the polypeptides,	
CC		and methods of identifying modulators of the polypeptides. The	
CC		polypeptides are useful for identifying modulator compounds which	

CC function as modulators, activators, repressors, agonists or antagonists
CC of the novel GPCR polypeptides including the GAL4 polypeptide. The
CC antibodies and nucleic acid probes as described above can be used to
CC detect the presence of the polypeptides and nucleic acids and are used to
CC diagnose a variety of diseases or disorders in which GPCRs are involved
CC e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma,
CC atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy,
CC chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease,
CC depression, epilepsy, macular degeneration, lymphoma, melanoma,
CC multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease,
CC psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis,
CC tuberculosis and many other diseases listed in the specification. The
CC probes and antibodies are also useful for diagnosing cognition and memory
CC disorders, anorexia, hormonal release disorders, cardiovascular activity
CC disorders, pain perception disorders, obesity, diabetes, Alzheimer's
CC disease. Preferably, compounds that decrease or increase
CC the expression of galanin receptor (GAL4) can be used to treat obesity,
CC diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is
CC useful for treating the above mentioned disorders by gene therapy
CC techniques. The present sequence is a novel GPCR polynucleotide of the
CC invention.

Sequence 1119 BP; 142 A; 420 C; 340 G; 217 T; 0 other;

Query Match 10.7%; Score 66; DB 24; Length 1119;
Best Local Similarity 50.2%; Pred. No. 0.00046;
Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

QY 112 gatgaccctctgctcatcgtggtggtccggcgccatgcttggaactggagctggg 171
II IIII II IIII IIII IIII IIII IIII IIII
DB 49 gactaccgacaccccccgcctgctgctgctgctgctgctgctgctgctgctgctg 108
II IIII IIII IIII IIII IIII IIII IIII IIII
QY 172 ctgacagcagcagctgagcgccctgcaatgttcatccgacgagcgagcgccctggcag 231
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 109 ctccctcctcaagcgctgagcct---ctgggtcttctctgctggcgcgctgctgctg 165
II IIII IIII IIII IIII IIII IIII IIII IIII
QY 232 gccctgcttctactctgttcaactgctgctgctgctgctgctgctgctgctgctgctg 291
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 166 gtgggtgagcgtgtacatgtgtaactgtgagcgagcagcgtgcttcaacctgctg 225
II IIII IIII IIII IIII IIII IIII IIII IIII
QY 292 cagctggtgctcaactctactccg-----gacctggccggagggcgctgctgacac 341
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 226 ccggttctgctctactctactgacacgacacactgacctccctccgacctctgtgcca 285
II IIII IIII IIII IIII IIII IIII IIII IIII
QY 342 gcggcgagcgacactactacgtgtccactatgcgagcggtgcttccgctgacatc 401
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 286 agcagcgagcgacatctccagatgaacatgtacgacgctgacatctcccgatgctc 345
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 402 agcgtgtgctgctgagcttctgtaacgctgacgagcgagcgagctgcccgtgccc 461
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 346 aacgtgagcagctacgacgacatcgtcacccgctgacacgctgacacgtgagcgccc 405
II IIII IIII IIII IIII IIII IIII IIII IIII
QY 462 cgggtgacgtgagbccccggcgagcgctgctgctgctgctgctgctgagcgctggcc 521
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 406 cgggtgagcgagcgtgctgctgagcggtggtggtggtggtggtggtggtggtggtg 465
II IIII IIII IIII IIII IIII IIII IIII IIII
QY 522 cctccctgctgagcgacagcgtg 545
II IIII IIII IIII IIII IIII IIII IIII IIII
DB 466 gcgcgcgcgtgacagcgccctg 489
II IIII IIII IIII IIII IIII IIII IIII IIII

RESULT 10
AAS98085/c
ID AAS98085 standard; DNA: 1237 BP.

XX AAS98085;

XX 12-MAR-2002 (first entry)

XX Human DNA for potential G protein-coupled receptor #42.

XX Human: G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;

KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW tuberculosis; cognition disorder; memory disorder; anorexia;
KW hormonal release disorder; cardiovascular activity disorder;
KW pain perception disorder; obesity; diabetes; obesity;
KW diabetes; hyperlipidaemia; stroke; gene therapy.

OS Homo sapiens.

PN WO200185791-A1.

PN 15-NOV-2001.

PF 11-MAY-2001; 2001WO-US15332.

XX 11-MAY-2000; 2000US-203217P.

PR 18-MAY-2000; 2000US-205945P.

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

XX Brown JP, Miller M, Burner G, Fabre-Suver C, Pritchard D;

XX WPI: 2002-066595/09.

PT Novel G protein-coupled receptor polypeptides including galanin

PT receptor polypeptides useful for identifying modulators that are useful

PT for treating Alzheimer's disease, psoriasis, melanoma, multiple

PT sclerosis, stroke

XX Claim 2; Page 72; 144pp; English.

CC The invention relates to an isolated polypeptide encoded by a
CC nucleic acid molecule that is at least 80% identical to the G
CC protein-coupled (GPCR) polynucleotides included in the specification.
CC Also included are probes based on the GPCR sequences (including
CC antisense probes), a host cell comprising an expression vector comprising
CC the GPCR sequence, antibodies raised against the polypeptides,
CC and methods of identifying modulators of the polypeptides. The
CC polypeptides are useful for identifying modulator compounds which
CC function as modulators, activators, repressors, agonists or antagonists
CC of the novel GPCR polypeptides including the GAL4 polypeptide. The
CC antibodies and nucleic acid probes as described above can be used to
CC detect the presence of the polypeptides and nucleic acids and are used to
CC diagnose a variety of diseases or disorders in which GPCRs are involved
CC e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma,
CC atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy,
CC chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease,
CC depression, epilepsy, macular degeneration, lymphoma, melanoma,
CC multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease,
CC psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis,
CC tuberculosis and many other diseases listed in the specification. The
CC probes and antibodies are also useful for diagnosing cognition and memory
CC disorders, anorexia, hormonal release disorders, cardiovascular activity
CC disorders, pain perception disorders, obesity, diabetes, Alzheimer's
CC disease. Preferably, compounds that decrease or increase
CC the expression of galanin receptor (GAL4) can be used to treat obesity,
CC diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is
CC useful for treating the above mentioned disorders by gene therapy
CC techniques. The present sequence is a novel GPCR polynucleotide of the
CC invention.

SQ Sequence 1237 BP; 243 A; 366 C; 463 G; 165 T; 0 other;

Query Match 10.7%; Score 66; DB 24; Length 1237;
Best Local Similarity 50.2%; Pred. No. 0.00046;
Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

QY 112 gatgaccctgctcatcgtggtggtccggcgccatgcttggaactggagctggg 171

Db 1086 GACTACGACCTACCAACGCGCTGACTGTGTGCTACACCTGTGCTGTGCGCGG 1027
Oy 172 ctgcagcaacgctgagcgccctggaatgtcatcgcagcgagcgccctggccag 231
Db 1026 CTCCTCCCAACGCGCTAGCCCT---CTGGCTCTCTGCGCGCTGCGCTGACACG 970
Oy 232 gccctgtctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 291
Db 969 GTGGTAGACGCTGACATGTGTAACCTGGCGGCGGACGACCTGCTCTTACCCCTCGGTG 910
Oy 292 cagctgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 341
Db 909 CCGCTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 850
Oy 342 ggcgcggggcgccactactactactactactactactactactactactactactc 401
Db 849 ACGACGGGCGCGCATCTTCCAGATGACGAGGACGCTGATCTGATGCTCATTC 790
Oy 402 agcgtgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 461
Db 789 AACGTGACCCCTACGCGCCCATGCTGACCCGCTGCGACCTGCGCACCTGCGGCGCC 730
Oy 462 cgtgtcctacgagbccccgcgcgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 521
Db 729 CGCGTGGGCGGCTGCTCTGCTGCGGCGGTGCGGCGGTCTCATCTGTGTTGGCGTCCC 670
Oy 522 cctccctcgtcgtgagacgcgtcg 545
Db 669 GCCGCCGCGTGCACAGGCCCTCG 646

RESULT 11

AAD08837
ID AAD08837 standard; cDNA; 1325 BP.

XX AAD08837;

DT 04-SEP-2001 (first entry)

DE Human G-protein coupled receptor-3 (GCR3-3) cDNA.

XX Human; G-protein coupled receptor-3; GCR3-3; gene therapy; cirrhosis;
KW transgenic animal; proliferative disorder; acinic keratosis; hepatitis
KW nephrotropic; cancer; breast; bladder; bone marrow; brain; uterus;
KW leukemia; adenocarcinoma; lymphoma; melanoma; myeloma; epilepsy; stroke;
KW neurological disorder; Alzheimer's disease; Parkinson's disease; nausea;
KW Huntington's disease; multiple sclerosis; dementia; angina pectoris;
KW central nervous system disorder; cardiovascular disorder; hypertension;
KW atherosclerosis; congestive heart failure; gastrointestinal disorder;
KW dysphagia; peptic esophagitis; spasm; gastritis; anorexia; pyrosis;
KW pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
KW inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
KW Addison's disease; allergy; asthma; diabetes mellitus; antithyroid;
KW atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
KW rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
KW metabolic disorder; obesity; nootropic; prozoacide; virucide; ss.

XX Homo sapiens.

FH Key Location/Qualifiers

FT CDS 42..1310

FT /tag- a

FT /product- "Human GCR3-3 protein"

XX WO200142288-A2.

XX 14-JUN-2001.

XX 07-DEC-2000; 2000WO-US33382.

XX 10-DEC-1999; 99US-0172852.

XX 22-DEC-1999; 99US-0171732.

PR 14-JAN-2000; 2000US-0176148.
PR 21-JAN-2000; 2000US-0177331.
XX
XX
XX (INCY-) INCYTE GENOMICS INC.
PI Burford N, Baughn MR, Au-Young J, Yang J, Lu DM, Reddy R;
XX WPI: 2001-381635/40.
DR P-FSDB: AAE04547.
XX
XX New human G-protein coupled receptor polypeptides for diagnosing,
PT preventing, and treating cell proliferative, neurological,
PT cardiovascular, gastrointestinal, autoimmune and metabolic disorders -
XX
PS Claim 5; Page 160-161; 175pp; English.

CC The present sequence is human G-protein coupled receptor-3 (GCR3-3)
CC cDNA. GCR3 is useful in somatic or germline gene therapy to correct a
CC genetic deficiency, to express a conditionally lethal gene product and
CC to express a protein which affords protection against intracellular
CC parasites and also for diagnosis of disorders associated with expression
CC of GCR3. GCR3 is also useful for generating hybridisation probes useful
CC in mapping the naturally occurring genomic sequences and to create
CC knockin humanised animals (pigs) or transgenic animals (mice or rats) to
CC model human diseases. GCR3 is used to diagnose, prevent and treat
CC proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
CC hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,
CC uterus cancer, leukemia, adenocarcinoma, lymphoma, melanoma and myeloma)
CC neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,
CC Parkinson's disease, multiple sclerosis, dementia and other central
CC nervous system disorders); cardiovascular disorders (angina pectoris,
CC hypertension, atherosclerosis, congestive heart failure);
CC gastrointestinal disorders (dysphagia, peptic esophagitis, esophageal
CC spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
CC pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/
CC inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
CC Addison's diseases, allergies, anaemia, asthma, diabetes mellitus, atopic
CC dermatitis, glomerulonephritis, Grave's disease, osteoarthritis,
CC psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
CC parasitic, protozoal and helminthic infections) and metabolic disorders
CC (obesity, osteoporosis, viral infections).

SQ Sequence 1325 BP; 195 A; 473 C; 403 G; 254 T; 0 other;

Query Match 10.7%; Score 66; DB 22; Length 1325;

Best Local Similarity 50.2%; Pred. NO. 0.00046;

Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

Oy 112 gatgacccctcgcgtctatcgtgtgtcgcgcgcgtcgtcgtcgtcgtcgtcgtcgtc 171
Db 90 gactaccgaccaccaccgcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 149
Oy 172 ctgcagcaacgctgagcgccctggaatgtcatcgcagcgagcgccctggccag 231
Db 150 ctcccccacacgagctagccct---ctgggtcttcctcgcgcgtcgtcgtcgtcgtcgtc 206
Oy 232 gccctgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 291
Db 207 gtgtgagcgtgtacatgtgtacacctgagcgagcagcctcctcaccctcgtcgt 266
Oy 292 cagctgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 341
Db 267 cccgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcgtcgt 326
Oy 342 ggcgcggggcgccactactactactactactactactactactactactactactc 401
Db 327 acgagcgcgccatcttcagatgacatgtacagcgagcgtgatcttcctcgtcctc 386
Oy 402 agcgtgtcgcgtcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 461
Db 387 aacgtgacgcgtacgcgcgcacgtcgtcgcgcgcgtcgtcgcgcacgtcgtcgtcgtcgtc 446

[illegible]

OY	342	ggcgccgagggcaccactactacgttgcaccactatctgcggcggttgcttcggccgagccatc	401
OY	402	agcggttcgccgttcggcttcgtctaaacggtttcccggggcccaaggcgagtctgcgcctgtgac	461
OY	462	cgggtgcctaaggbgccccgcgcgcgtctgccttcggccttcgttcgttcgttcggcggttcgac	521
DB	687	cgcgtgcgcgcgctcgtctcgtcctgcgttggtggtggcgctcatcctgttgttccgtgccc	746
OY	522	cctccctgccttgagcacgcctg	545
OY	747	gcgcgcgcgtgcacagcctctgg	770
 RESULT_14 AAD26369			
ID	AAD26369	standard; CDNA; 2444 BP.	
XX	AAD26369;		
DT	26-MAR-2002	(first entry)	
DE	Human G-protein coupled receptor 1 (GCRRC-1) cDNA.		
KW	Human; G-protein coupled receptor 1; cell proliferative disorder;		
KW	arteriosclerosis; hepatitis; cancer; neurological disorder; epilepsy;		
KW	Alzheimer's disease; Parkinson's disease; cardiovascular disorder;		
KW	atherosclerosis; hypertension; myocardial infarction; peptic ulcer;		
KW	gastrointestinal disorder; dysphagia; anorexia; autoimmune disorder;		
KW	acquired immune deficiency syndrome; inflammatory disorder; infection;		
KW	Addison's disease; allergy; Grave's disease; metabolic disorder; AIDS;		
KW	diabetes; obesity; osteoporosis; gene therapy; GCRRC-1; ss.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	CDS	350..1468	
FT		/tag= a	
XX		/product= "Human GCRRC-1 protein"	
PN	WO200187937-A2.		
PD	22-NOV-2001.		
PF	17-MAY-2001; 2001WO-US16285.		
PR	18-MAY-2000; 2000US-205628P.		
PR	22-MAY-2000; 2000US-206222P.		
PR	25-MAY-2000; 2000US-207566P.		
PR	02-JUN-2000; 2000US-208834P.		
PR	02-JUN-2000; 2000US-208861P.		
PA	(INCY-) INCYTE GENOMICS INC.		
P1	Patterson C, Lu DAM, Thornton M, Lu Y, Tribouley CM, Graul R;		
P1	Khan FA, Gandhi AR, Walla NK, Nguyen DB, Yue H, Hafalia A;		
P1	Elliott VS, Lal P, Reddy R, Kallikch DA, Tang TY, Au-Young J;		
DR	WPI; 2002-089844/12.		
DR	P-PSDB; AME16170.		
PT	Novel G-protein coupled receptors and polynucleotides useful for		
PT	diagnosis, treatment and prevention of disorders of cell proliferation		
PT	neurological, cardiovascular, metabolic disorders and viral infections		
PS	Claim 5; Page 110-111; 11pp; English.		
CC	The invention relates to human G-protein coupled receptor (GCRRC)		


```

Db      90  gactaccgactaccacacgactgactggtggtctacaagcttggtgctgctgcgcgg  149
QY      172  ctgacagacacagctgagcgccctgagcaatgttcatccgacagcgcgccctgagccag  231
Db      150  ctccccctcaacgctagacct--ctgggtcttccctgcgcgcgtgctgcaactcg  206
QY      232  gccctgctctctacactgttcaacctgtgctgtgtagttcttcaagctcaagctg  291
Db      207  gtggtgagcgtgtacatgtttaacctggcgacagcagacctgtctcttcaacctctgctg  266
QY      292  cagctgtgctcaactaactg-----ggcctggcccgagggccgcctgcac  341
Db      267  ccggtctctctctactaactgacactgacacacatgccccttcccgacctctgtgccaag  326
QY      342  gcggccggggccacactaactaactgttcaactatgcggcggtgtgtcttgcgcgcatac  401
Db      327  acgacgggcccacatcttccagatgaacatgtlacggcagctgcaatcttctgatatctac  386
QY      402  agcgtgacgctgagcgtctgtagcggttcccgagcccaaggcggtcgcgcgcctgccc  461
Db      387  aacgtgacacgctacgcgcgcacatctgtgaacccgctgacactgcgcacactgcgcgccc  446
QY      462  cgggtgctaacggbgccccgcgcgcgtgcctgcgccttgcctgctggcctggcgccctggcc  521
Db      447  cgcgtagcgcgctgctgctgctggtggtggtgcgtcaccctcgtgtgttgcgtgccc  506
QY      522  cctccctgacctggagacacgctgg  545
Db      507  gcgcgcgcgtgacagagccctcg  530

```

Search completed: July 13, 2002, 02:11:42
 Job time: 4365 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 21:59:49 ; Search time 1695.85 Seconds
(without alignments)
4918.547 Million cell updates/sec

Title: US-09-899-513-1

Perfect score: 618
Sequence: 1 atgggggctgccttgcgc.....ggctgcgcgcacacgtga 618

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vtc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	482.2	78.0	483	9	AI935035 wd17g08.x
2	469.8	76.0	490	10	BF195670 BF195670.7n86906.x
3	458.6	74.2	466	9	AI277080 q140h02.x
4	456.6	73.9	440	9	AI264420 qk54f04.x
5	446.6	72.3	469	10	BF000091 BF000091.7n17d11.x
6	389.6	63.0	392	10	BF108800 BF108800.7n152h02.x
7	385	62.3	432	9	AM072531 xa08e08.x
8	303.6	49.1	333	9	AM014139 UI-H-B10-
9	200	32.4	220	9	AM014455 UI-H-B10-
10	181.2	29.3	1083	10	BF307868 BF307868.601890683
11	142.6	23.1	189	9	AM138178 UI-H-B11-
12	72.6	11.7	925	12	AL055013 drosoph11
13	68.4	11.1	935	12	AL060051 drosoph11
14	67.2	10.9	935	12	AL066051 drosoph11
15	66.4	10.7	932	12	AL066742 drosoph11
16	65.6	10.6	816	10	BG756157 BG756157.602713458
17	65.2	10.6	117	9	AM136981 UI-H-B11-

C 18	64.4	10.4	932	12	CNS00720	AL066742 drosoph11
C 19	64	10.3	1101	12	CNS0175X	AL108460 drosoph11
C 20	63.8	10.3	925	12	CNS0091P	AL055013 drosoph11
C 21	63.2	10.2	961	12	AG041025	AG041025 Pan trogl
C 22	63.2	10.2	1189	12	AG032118	AG032118 Pan trogl
C 23	62.8	10.2	1203	12	CNS01574	AL106054 drosoph11
C 24	62.2	10.1	1233	12	AG074706	AG074706 Pan trogl
C 25	61.6	10.0	914	10	BF663176	BF663176 602144305
C 26	61.6	10.0	1091	12	AG060140	AG060140 Pan trogl
C 27	61.4	9.9	567	12	CNS075AL	AL429856 clone BA0
C 28	61.4	9.9	990	12	AG043619	AG043619 Pan trogl
C 29	61.2	9.9	334	10	BM491572	BM491572 pyp2n.pk0
C 30	61	9.9	844	12	CNS0052P	AL056652 drosoph11
C 31	60.8	9.8	846	10	BI951973	BI951973 HVSmem000
C 32	60.2	9.7	862	12	AG036151	AG036151 Pan trogl
C 33	60	9.7	1040	12	AG036367	AG036367 Pan trogl
C 34	59.8	9.7	784	12	AG161654	AG161654 Pan trogl
C 35	59.8	9.7	1221	12	AG058113	AG058113 Pan trogl
C 36	59.6	9.6	986	12	AG074680	AG074680 Pan trogl
C 37	59.4	9.6	1014	10	BF975186	BF975186 602244682
C 38	59.2	9.6	839	12	CNS004NB	AL054280 drosoph11
C 39	59.2	9.6	1148	10	BI834824	BI834824 603090273
C 40	59	9.5	776	12	CNS010RY	AL093532 drosoph11
C 41	59	9.5	1100	12	CNS016KD	AL106855 drosoph11
C 42	58.8	9.5	772	12	AQ740673	AQ740673 HS-5507_A
C 43	58.8	9.5	1281	10	BG852363	BG852363 102403440
C 44	58.8	9.5	1350	10	BM450237	BM450237 ACNCOURT
C 45	58	9.4	814	9	BE055163	BE055163 GA_Ea003

ALIGNMENTS

RESULT 1
AI935035/c 483 bp mRNA linear EST 17-DEC-1999
LOCUS wd17g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2328446 3', mRNA sequence.
ACCESSION AI935035
VERSION AI935035.1 GI:5673905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 483)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 894 Std Error: 0.00
Seq primer: -40bp from Gibco
High quality sequence stop: 442.
Location/Qualifiers
1. 483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2328446"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH119W, testis NHT, and B-cell
NCL-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 466)	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	NCI-CCAP http://www.ncbi.nlm.nih.gov/ccicgap .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			

Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 466)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA library preparation: M. Bento Soares, Ph.D.
cDNA library arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdnp/image/image.html
Insert length: 978 Std Error: 0.00
Seq primer: -40UP from Gdbco
High quality sequence stop: 431.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1874835"
/clone_id="NCI_CGAP C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bernaldo. "
BASE COUNT
      89 a      153 c      177 g      45 t      2 others
ORIGIN

```

Query Match	74.2%	Score 458.6;	DB 9;	Length 466;
Best Local Similarity	99.1%	Pred. No. 1.9e-69;		
Matches 459; Conservative	1;	Mismatches 3;	Indels 0;	Gaps 0

[illegible]

Db 46 GGGCCCTCCCTGCCTGAAGCACCGCTGGCAAGCTCGGGCTG 4

RESULT	4		
LOCUS	A1264420/c	460 bp	mRNA
DEFINITION	linear EST 28-JAN-1999		
ACCESSION	A1264420		
VERSION	G554f04.x1		
KEYWORDS	NCI CGAP CoB Homo sapiens cDNA clone IMAGE:1872799 3'		
SOURCE	similar to SM:PXY4_HUMAN P51562 URIDINE NUCLEOTIDE RECEPTOR ;,		mRNAS
ORGANISM	sequence.		
	A1264420		
	A1264420.1	GI:3872623	
	EST.		
	human.		
	Homo sapiens		

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 460)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

JOURNAL
COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbfp/image/image.html
Insert length: 1000 Std Error: 0.00
Seq primer: -40UP from glbco
High quality sequence stop: 449.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1872799"
/clone_1id="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT
ORIGIN
87 a 151 c 179 g 42 t 1 others

```

Query Match	73.9%	Score 456.6	DB 9	Length 460
Best Local Similarity	99.3%	Pred. No. 4.1e-69		
Matches 457; Conservative	1;	Mismatches 2;	Indels 0;	Gaps 0;

QY	112	gataacctgctgcttcacatcgatgagcccgagccatgacctggacgagccctggag	171
Db	460	gataaccctgctccttcacatgctgctccgagccctgacgttgagacgtggcctggcg	401
QY	172	ctgcgacgccaacgctgagcgagccctgcaatltcacccgacgagcgagcgctctgggcca	231
Db	400	ctgcgacgccaacgctgagcgagccctgcaatlttcacccgacgagcgagcgctctgggcca	341
QY	232	gcccctgcttcttaacctggttcaacctgagctctggttgatgagatctctcaagctcaagctg	291
Db	340	gcccctgcttcttcttaacctggttcaacctgagctctggttgatgagatctctcaagctcaagctg	281
QY	292	caagctgagcttaacctactactactgagcgctgcccggagagcgccgctgcgaagcgagccggag	351
Db	280	caagctgagcgcttaacctactactactgagcgctgcccggagagcgccgctgcgaagcgagccggag	221

```

Oy 352 ccaccactacgtgtccacatgatcgagtggtcttcgcgcgcgtcatcaagcgtgtgcc 411
    |||||||
Db 220 CCACCTACTACGCTGTCCACCTATGCGCGGTGTCTTCGCCGCGCATCAGGCTGTGCC 161
    |||||||
Oy 412 gctcgagcttcgtaagcgtgtccgagcgaagcgtgcgcgcgtgcgcgcgtgtgcctac 471
    |||||||
Db 160 GCTGCGGCTTCGTAACGCTGTCCGCGCGCATCAGGCTGTGCCGCGCATCAGGCTGTGCC 101
    |||||||
Oy 472 gpbagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 531
    |||||||
Db 100 GGGGCCCCGGCGCGCTGCGCTGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 41
    |||||||
Oy 532 tggagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 571
    |||||||
Db 40 TGAAGCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1
    |||||||

RESULT 5
BFO00091/c 449 bp mRNA linear EST 06-OCT-2000
INITIATION 7h17d11 x1 NCI CGAP Col6 Homo sapiens CDNA clone IMAGE:3316245 3'
    similar to contains PTRS.t3 TAR1 repetitive element ;, mRNA
    sequence.
ACCESSION BFO00091 GI:10700366
VERSION BFO00091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgaps-r@mail.nih.gov
    Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
    Ph.D.
    ' CDNA Library Preparation: M. Bento Soares, Ph.D.
    CDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LINL, send email to:
    info@image.lnl.gov
    Seq primer: -40UP from Gibco
    High quality sequence stop: 388.
FEATURES
    source
    Location/Qualifiers
    1..449
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_1b="IMAGE:3316245"
    /clone_1lb="NCI CGAP Col6"
    /tissue_type="colon tumor, RER+"
    /lab_host="DH10B"
    /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
    modified polylinker; Site_1: Not I; Site_2: Eco RI;
    Plasmid DNA from the normalized library NCI CGAP Col6 was
    prepared, and ss circles were made in vitro. Following HAP
    purification, this DNA was used as tracer in a subtractive
    hybridization reaction. The driver was PCR-amplified cDNAs
    from a pool of 5,000 clones made from the same library
    (cloneids 1057416-1061255, and 1144584-1145351).
    Subtraction by Bento Soares and M. Patricia Bonaldo. "
BASE COUNT 87 a 150 c 172 g 40 t
ORIGIN

```

```

Query Match 72.3%; Score 446.6; DB 10; Length 449;
Best Local Similarity 99.6%; Pred. No. 2,1e-67;
Matches 447; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 98 gctcgagcctctgagatgacgcgtctcgcttcacgtgtgtgtccgcgcgcgtatgcttgg 157
    |||||||
Db 449 GCTGCAAGCCTCTTGAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 390
    |||||||
Oy 158 cacttgagcctctgagatgacgcgtctcgcttcacgtgtgtgtccgcgcgcgtatgcttgg 217
    |||||||
Db 389 CACTTGAGCCTCTTGAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 330
    |||||||
Oy 218 ggcgcctctgagatgacgcgtctcgcttcacgtgtgtgtccgcgcgcgtatgcttgg 277
    |||||||
Db 329 GGGGCGTGGGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 270
    |||||||
Oy 278 tcaagctacgcgtcgcgcgtctgcttcacgtgtgtgtccgcgcgcgtatgcttgg 337
    |||||||
Db 269 TCAAGCTACGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210
    |||||||
Oy 338 ccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 397
    |||||||
Db 209 CCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150
    |||||||
Oy 398 cactcgagctctgagatgacgcgtctcgcttcacgtgtgtgtccgcgcgcgtatgcttgg 457
    |||||||
Db 149 CATCAGCGTGGGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90
    |||||||
Oy 458 ggcgcgcgtctgagatgacgcgtctcgcttcacgtgtgtgtccgcgcgcgtatgcttgg 517
    |||||||
Db 89 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30
    |||||||
Oy 518 ggcgcgcgtctgagatgacgcgtctcgcttcacgtgtgtgtccgcgcgcgtatgcttgg 546
    |||||||
Db 29 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1
    |||||||

RESULT 6
BFI08800/c 392 bp mRNA linear EST 20-OCT-2000
DEFINITION 7152h02 x1 Soares_NSF_F8_9M_OT_PA_P_S1 Homo sapiens CDNA clone
    IMAGE:3525338 3', mRNA sequence.
ACCESSION BFI08800
VERSION BFI08800.1 GI:10938490
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 392)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgaps-r@mail.nih.gov
    This clone is available royalty-free through LINL; contact the
    IMAGE Consortium (info@image.lnl.gov) for further information.
    Seq primer: -40UP from Gibco
    High quality sequence stop: 358.
FEATURES
    source
    Location/Qualifiers
    1..392
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_1b="Soares_NSF_F8_9M_OT_PA_P_S1"
    /lab_host="DH10B"
    /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
    a modified polylinker; Site_1: Not I; Site_2: Eco RI;
    Equal amounts of plasmid DNA from five normalized
    libraries were mixed, and ss circles were made in vitro.
    Following HAP purification, this DNA was used as tracer in
    a subtractive hybridization reaction. The driver was
    PCR-amplified cDNAs from pools of 5,000 clones made from
    the same 5 libraries. The pools consisted of the following
    libraries and cloneids: Soares NBHSF pool 1:
    309384-310919, 323208-325895 Soares NB2HP pool 1:

```


145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares_NFL_T_GBC_S1
758280-760583, 772104-774407 Soares_NbHPA pool 1:
304776-306311, 320136-322823, 326280-326653 Soares_NbHOT
pool 1: 723720-726407, 73980-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo.*

BASE COUNT 72 a 131 c 153 g 36 t

Query Match 63.0%; Score 389.6; DB 10; Length 392;
Best Local Similarity 99.5%; Pred. No. 1.2e-57;
Matches 390; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

177 agccaacgtggcggccgggaatgttaccgagcgcgcgccggcgccgacct 236
|||||
392 agccaacgtggcggccgggaatgttaccgagcgcgcgccggcgccgacct 333

237 gctctctactgttcaacgtgctgtgtatgaatgttctcaacgtcagctgcaagt 296
|||||
332 gcttctactgttcaacgtgctgtgtatgaatgttctcaacgtcagctgcaagt 273

297 gtggtcaactactactgagcgtgccccgagagcgccgtgcaacgccccgggacac 356
|||||
272 gtggtcaactactactgagcgtgccccgagagcgccgtgcaacgccccgggacac 213

357 tactaggtgtcaactactgagcgtgccccgagagcgccgtgcaacgccccgggacac 416
|||||
212 tactaggtgtcaactactgagcgtgccccgagagcgccgtgcaacgccccgggacac 153

417 ggtcttcaacgtgctgccccgagagcgccgtgcaacgccccgggacacgtgcaacg 476
|||||
152 ggtcttcaacgtgctgccccgagagcgccgtgcaacgccccgggacacgtgcaacg 93

477 ccgagcgagcgtgctgccccgagagcgccgtgcaacgccccgggacacgtgcaacg 536
|||||
92 cccgcgtgagcgtgctgccccgagagcgccgtgcaacgccccgggacacgtgcaacg 33

537 cagcgtggagagcgtgccccgagagcgccgtgcaacgccccgggacacgtgcaacg 568
|||||
32 cagcgtggagagcgtgccccgagagcgccgtgcaacgccccgggacacgtgcaacg 1

RESULT 7 432 bp mRNA linear EST 20-OCT-2000
AM072531 LOCUS

DEFINITION x08e08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2567750 3 similar to contains PTRS.b2 TAR1 repetitive
element; , mRNA sequence.

ACCESSION AM072531
VERSION AM072531.1 GI:6027529
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 432)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

FEATURES
source location/Qualifiers
1..432
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="IMAGE:2567750"
/clone.lib="Soares_NFL_T_GBC_S1"
/lab_host="pDH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not 1; Site:2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-GAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 84 a 145 c 164 g 38 t 1 others

Query Match 62.3%; Score 385; DB 9; Length 432;
Best Local Similarity 95.4%; Pred. No. 7.6e-57;
Matches 418; Conservative 1; Mismatches 13; Indels 6; Gaps 2;

127 ttcaatgtgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgtg 186
|||||
432 ttcaatgtgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgtg 373

187 gcgagcctggaatgttcaacgtgccccgagcgtgctgtgtatgaatgttctcaacgt 246
|||||
372 gcgagcctggaatgttcaacgtgccccgagcgtgctgtgtatgaatgttctcaacgt 313

247 ctgtcaacgtgctgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgt 306
|||||
312 ctgtcaacgtgctgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgt 253

307 tactactgagcgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgt 366
|||||
252 tactactgagcgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgt 198

367 ccaactatgagcgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgt 426
|||||
197 ccaactatgagcgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgt 139

427 gcgagcctggaatgttcaacgtgccccgagcgtgctgtgtatgaatgttctcaacgt 486
|||||
138 gcgagcctggaatgttcaacgtgccccgagcgtgctgtgtatgaatgttctcaacgt 79

487 ctgctgagcgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgtg 546
|||||
78 ctgctgagcgtgccccgagcgtgctgtgtatgaatgttctcaacgtcagctgcaacgtg 19

547 caagctcgagcgtgctgct 564.
|||||
18 caagctcgagcgtgctgct 1

RESULT 8 333 bp mRNA linear EST 10-SEP-1999
AM014139 LOCUS

DEFINITION UI-H-B10-aa1-b-07-0-UI.s1 NCI-CCAP_Sub1 Homo sapiens cDNA clone
IMAGE:2709445 3', mRNA sequence.

ACCESSION AM014139
VERSION AM014139.1 GI:5862896
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 333)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Oligo-dt track not found. Not I site shown in beginning of sequence
 is likely internal to the message. CDNA library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html The following repetitive
 elements were found in this cDNA sequence: 236-273,
 >GC-richlow-complexity
 Seq primer: M13 Forward
 POLYA-NO.

FEATURES

Location/Qualifiers

```

1..333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2709445"
/clone_lib="NCI-CGAP_Sub1"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI-CGAP_Sub1 library is a subtracted library derived from
BI. BI constitutes a mixture of 21 normalized or
subtracted NCI-CGAP libraries: NCI-CGAP_C04,
NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_C010, NCI-CGAP_C016,
NCI-CGAP_K105, NCI-CGAP_K102, NCI-CGAP_K103,
NCI-CGAP_K101, NCI-CGAP_Lym2, NCI-CGAP_Br2, NCI-CGAP_C08,
NCI-CGAP_C101, NCI-CGAP_Le12, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
NCI-CGAP_Lu24, NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with a
driver whose composition is detailed below: NCI-CGAP_K103
pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
Clones 1322376-1323911, 1456008-1456775, 1500552-1502855)
) NCI-CGAP_K105 pool 1 LAM 3338-3342, 3722-3725,
3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI-CGAP_Lu5 pool 1 LAM
3575-3582, 3851-3854 (IMAGE Clones 141920-1417991,
1520904-1522439) NCI-CGAP_GC4 pool 1 LAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI-CGAP_Pr22 pool 1
LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 1217928-1220615)
NCI-CGAP_C010 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351) The resulting
subtracted library contained 530,000 recombinants.
Subtraction was performed as previously described [Bonaldo
Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB=NCI-CGAP_Br2
TAG_TISSUE=breast
TAG_SEQ=AAACC"
60 a 125 g 33 t

```

BASE COUNT
 ORIGIN

Query Match 49.1%; Score 303.6; DB 9; Length 333;
 Best Local Similarity 96.1%; Pred. No. 6.5e-43;

Matches 321; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

```

OY 255 cctgctctgtgtatgagttctctacgtcaagctgtgtgtctactactact 314
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 333 CTTGGCTATGGTGTATGATGATATTCATGCTCAACCTGCACGCTGCTACATCACTACT 274
OY 315 gggcctgtcccgagggcgcctcgcacggcgggcgggcgacactactactgttccactat 374
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 273 GGGCGGTGGCCGGGGGGCGCCT-CCACGGCGCCGGGGCGCAATATATGACCACTAT 215
OY 375 ggcgcggtgtctctcgcgcgtcatcagcggtgtcgcgcgtcgcgttcgtacgcgttccc 434
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 214 GCGGCGGTGTCTTCCGCGCGCTCATCAAGCTGTGCGCGCTTCTGTACGCGGCTCC 155

```

FEATURES

source

JOURNAL

COMMENT

ORGANISM

REFERENCE

AUTHORS

TITLE

UNPUBLISHED (1997)

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Oligo-dt track not found. Not I site shown in beginning of sequence
 is likely internal to the message. CDNA library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html
 Seq primer: M13 Forward
 POLYA-NO.

FEATURES

source

```

1..220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2709396"
/clone_lib="NCI-CGAP_Sub1"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI-CGAP_Sub1 library is a subtracted library derived from
BI. BI constitutes a mixture of 21 normalized or
subtracted NCI-CGAP libraries: NCI-CGAP_C04,
NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_C010, NCI-CGAP_C016,
NCI-CGAP_K105, NCI-CGAP_K102, NCI-CGAP_K103,
NCI-CGAP_K101, NCI-CGAP_Lym2, NCI-CGAP_Br2, NCI-CGAP_C08,
NCI-CGAP_C101, NCI-CGAP_Le12, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
NCI-CGAP_Lu24, NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with a
driver whose composition is detailed below: NCI-CGAP_K103
pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
Clones 1322376-1323911, 1456008-1456775, 1500552-1502855)
) NCI-CGAP_K105 pool 1 LAM 3338-3342, 3722-3725,
3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI-CGAP_Lu5 pool 1 LAM
3575-3582, 3851-3854 (IMAGE Clones 141920-1417991,
1520904-1522439) NCI-CGAP_GC4 pool 1 LAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI-CGAP_Pr22 pool 1
LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 1217928-1220615)
NCI-CGAP_C010 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351) The resulting

```


RESULT	12	
CNS0091P		
LOCUS	CNS0091P	925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19b16 of RPCT-98 library from Drosophila melanogaster (fruit fly) genomic survey sequence.	
ACCESSION	AL053013	
VERSION	AL053013.1	GI:4934461
KEYWORDS	GSS.	

RESULT	13				
CNS006XK/c					
LOCUS	CNS006XK	935 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #				
	BARIN09 of Rp11.98 library from Drosophila melanogaster (fruit				
	fly), genomic survey sequence.				
ACCESSION	AL066051				

RESULT	15	
LOCUS	CNS00720	
DEFINITION	CNS00720. 932 bp DNA linear GSS 03-JUN-1999	
ACCESSION	Drosophila melanogaster genome survey sequence T7 end of Bac #	
VERSION	BACR1B09 of RPc1-98 library from Drosophila melanogaster (fruit	
KEYWORDS	fly), genomic survey sequence.	
SOURCE	AL066742	
ORGANISM	AL066742.1 GI:4945205	
	GSS.	
	fruit fly.	
	Drosophila melanogaster	
	Euryptera; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
	1 (bases 1 to 932)	
REFERENCE	Genoscope.	
AUTHORS	Direct Submission	
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :	
JOURNAL		

Mon Jul 15 14:32:24 2002

us-09-899-513-1.rst

Page 11

OM of: US-09-899-513-2 to: GenEmbl:* out_format : pts

Date: Jul 13, 2002 3:10 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=fasta.p2n.model -DEV=xlh
-Q/cgn2.1/USFT0_spool/US0989513/runatc_12072002.082700.12722/app.query.fasta_1.264
-DB=GenEmbl -QFT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=1.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -OGAPOP=10.000 -OGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YCAPOP=10.000 -YGAPEXT=0.500
-DELOR=6.000 -DELEXT=7.000 -SPART=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pts
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0989513.cgn1.1.3962 -NCPU=6 -ICPU=3 -LONGLOG
-LEV_TIMEOUT=120 -NARN_TIMEOUT=30 -NO_XLPMX -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-899-513-2
Query Length: 205
Database: GenEmbl:*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1815.170000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb.htg:AL354876	- 1119.00	627.00	1.5e-26	152947	AL354876 Homo sapiens chrom
gb.htg:AL607089	- 1119.00	625.56	1.8e-26	152947	AL607089 Homo sapiens chrom
gb.pr:AF080214	+ 177.50	122.23	191.17	1534	AF080214 Homo sapiens protease
gb.pat:AR108663	+ 177.50	115.25	468.09	4895	AR108663 Sequence 1 from paten
gb.pr:AF055917	+ 177.50	115.25	468.09	4895	AF055917 Homo sapiens protease
gb.pat:AX211769	+ 177.50	115.21	470.30	4925	AX211769 Sequence 1 from Paten
gb.pr:AF384819	+ 177.50	109.94	924.76	11828	AF384819 Homo sapiens coagula
gb.htg:AC008737	- 171.50	91.61	9.7e+03	248680	AC008737 Homo sapiens chrom
gb.pr:HSU91939	- 171.50	119.10	285.46	1400	U91939 Human putative G protei
gb.pr:HUMCHRA	- 168.50	115.97	426.64	159361	J03483 Human chromogranin A mR
gb.htg:AC027113	+ 167.50	88.63	1.4e+04	159361	AC027113 Homo sapiens chrom
gb.pr:AC098756	+ 167.50	88.34	1.5e+04	167302	AC098756 Homo sapiens chrom
gb.pr:HUMCHRA	- 164.50	113.88	557.93	1811	J03915 Human chromogranin A mR
gb.pr:BC012755	- 163.50	114.73	500.47	1432	BC012755 Homo sapiens, clone M
gb.pr:BC009384	- 163.50	113.79	564.57	1674	BC009384 Homo sapiens, Similar
gb.ov:G6U59421	+ 162.50	113.54	582.54	2008	BC006459 Homo sapiens, chromog
gb.pat:AR096465	+ 162.50	112.69	649.66	2008	BC006459 Homo sapiens, chromog
gb.pr:BC006459	+ 163.00	112.69	649.66	2008	BC006459 Homo sapiens, chromog
gb.pr:AF100206	+ 163.00	115.77	437.57	1148	AF100206 Macaca mulatta paten
gb.pat:AR096464	+ 162.50	116.60	393.32	954	AR096464 Sequence 5 from paten
gb.pat:AR096466	+ 162.50	114.53	513.30	1347	AR096466 Sequence 8 from paten
gb.ov:G6U59421	+ 162.50	113.54	582.54	1587	U919421 Gallus gallus histidin
gb.pat:AR096465	+ 162.50	113.54	582.54	1587	AR096465 Sequence 6 from paten
gb.pat:AR096463	+ 162.50	111.49	758.20	2233	AR096463 Sequence 4 from paten
gb.pat:AR096462	+ 162.50	106.60	1.4e+03	5027	AR096462 Sequence 2 from paten
gb.pat:AX068253	+ 162.00	115.38	460.33	1116	AX068253 Sequence 15 from Paten
gb.pat:AX076182	+ 162.00	115.36	461.28	1119	AX076182 Sequence 1 from Paten
gb.pat:AX138829	+ 162.00	115.36	461.28	1119	AX138829 Sequence 1 from Paten
gb.pr:HSAT27207	+ 162.00	115.36	461.28	1119	AF172207 Homo sapiens mRNA for
gb.pr:AF411112	+ 162.00	115.36	461.28	1119	AF411112 Homo sapiens G protei
gb.pat:AX168124	+ 162.00	114.35	525.54	1325	AX168124 Sequence 42 from Paten
gb.pat:AX352473	+ 162.00	113.35	525.54	1325	AX352473 Sequence 25 from Paten
gb.pat:AX068253	+ 162.00	112.78	642.76	1720	AX068253 Sequence 25 from Paten
gb.pat:AX305129	+ 162.00	110.66	842.94	2444	AX305129 Sequence 9 from Paten
gb.pat:AX128511	+ 162.00	110.57	852.51	2480	AX128511 Sequence 19 from Paten
gb.htg:AC079387	- 162.00	82.91	3.0e+04	245880	AC079387 Homo sapiens chrom
gb.ba:SCF55	- 161.50	93.80	7.3e+03	38426	AL132991 Streptomyces coelic
gb.pr:HSU03642	+ 160.00	112.14	697.00	1583	U03642 Human G protein-coupled
gb.pr:HSU03642	+ 160.00	111.37	769.98	1801	HSU03642 Human G protein-coupled
gb.pat:AR107256	+ 160.00	111.13	793.30	1872	AR107256 Sequence 39 from Paten
gb.pr:AP001786	- 160.00	83.94	2.6e+04	17144	AP001786 Homo sapiens genom
gb.pr:BC001059	- 159.50	110.78	830.05	1894	BC001059 Homo sapiens, chromog

```
gb.ba:AF222792 - 159.50 103.92 2.0e+03 5921 AF222792 Streptomyces sp. C
gb.pat:E37240 + 159.00 113.55 581.71 1140 E37240 Novel physiological
gb.pat:E63120 + 159.00 113.55 581.71 1140 E63120 Peptide derivative.
gb.pr:HUMSOWATA + 159.00 112.58 658.99 1340 L07833 Homo sapiens somatos
gb.pl:AP003418 - 159.00 84.38 2.5e+04 145265 AP003418 Oryza sativa gen
```

seq_name: gb.htg:AL354876

seq_documentation_block:

```
LOCUS AL354876 152947 bp DNA linear HTG 27-JUN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-244H3, *** SEQUENCING IN
PROGRESS ***, 3 unordered pieces.
```

```
ACCESSION AL354876
VERSION AL354876.9 GI:14575158
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
```

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jun 28, 2001 this sequence version replaced gi:13234889.

Center: Sanger Centre
Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

Project Information

Center project name: Ba244H3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 152070 bases at least Q40
Consensus quality: 152318 bases at least Q30
Consensus quality: 152484 bases at least Q20
Insert size: 152747; sum-of-contigs
Insert size: 138665; 8.5% error; agarose-fp
Quality coverage: 9.58x in Q20 bases; sum-of-contigs Quality
coverage: 10.56x in Q20 bases; agarose-fp

***** NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 54502: contig of 54502 bp in length
* 54503 54602: gap of 100 bp
* 54603 99405: contig of 44803 bp in length
* 99406 99505: gap of 100 bp
* 99506 152947: contig of 53442 bp in length.

Location/Qualifiers

```
1. 152947
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-244H3"
/clone_id="RP11-11.1"
1. 54502
/note="assembly fragment:02081
fragment chain:1"
```

misc.feature

```
54603..99405
/note="assembly fragment:01111
fragment chain:1"
99506..152947
/note="assembly fragment:03694"
```

BASE COUNT 46288 a 33896 c 32934 g 39628 t 201 others
ORIGIN

alignment_scores:
Quality: 1119.00 Length: 205
Ratio: 5.459 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-899-513-2 x AL354876/rev ..

Align seg 1/1 to reverse of: AL354876 from: 1 to: 152947

```

1 MetGlyAlaAlaPheValAlaSerLeuArgSerAsnLeuSerSerAlaThr 17
|||||
54085 ATGGGGGCTGCTTGTGCTAGCCTCCGACAGTAACTTTCTTCTGCGCAC 54036
17 rSerArgSerGluMetAsnSerSerValGlyAspLeuGlyValGlyGlyC 34
|||||
54035 TTCAAGCTCAGACATGAACAGCACTGTTGGGAGCTGGTGTGGCGGCT 53986
34 ySserLeuTPAspAspProAlaArgPheIleValProAlaAlaTyr 50
|||||
53985 GCAAGCTCTGGATGACCTGCTGCTTCATCGTGGTGCGCGGCTCAT 53936
51 AlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMe 67
|||||
53935 GCCTTGACACAGGCGCTGGGCTCCAGCCAACTGGGCGGCTGGGCAT 53886
67 tPheIleArgSerGlyValArgLeuGlyGlnAlaLeuLeuLeuTyrLeuP 84
|||||
53885 GTTCATATCCGACGCGGCGGCGCTGGGCGCAGCCCTGCTCTACCTGT 53836
84 heAsnLeuAlaLeuValAlaSpGluPhePheThrLeuThrLeuGlnLeuTyr 100
|||||
53835 TCAACCTGCTGCTGTGATGATGATTCTTACAGCTCAGCTGACGCTGTGG 53786
101 LeuThrTyrTyrLeuGlyLeuAlaArgArgProProAlaThrArgProG 117
|||||
53785 CTCACCTACACTACCTGGGCGCTGGCGGAGCGCGCTGGCACGCGCGGG 53736
117 yProProThrThrcysProProMetArgArgTyrSerSerProArgSers 134
|||||
53735 GCCACCTACACTACGTCATCACCCTATGCGGCGGTGCTTCCGCGGCTCAT 53686
134 eAlaGlyAlaAlaAlaAlaSerTyrAlaValProGlyProGlyArgLeu 150
|||||
53685 CAGCGTGTGCGGCTGCGGCTTCGTACGCGGTCCCGGCGGCGGCGGCTG 53636
151 ProAlaTrpProGlyAlaTyrGlyAlaProArgAlaLeuProAlaProSe 167
|||||
53635 CCGGCTGAGCGCGGCTGCTACGCGGCGCGCGGCGGCTGGCGCTTC 53586
167 rProGlyTyrArgAlaTrpProLeuProAlaTrpSerThrAlaGlyGlnA 184
|||||
53585 GCCGTGCTGGGCGGCTGGCGGCTTCCTGCTGAGACACCCCTGGCGCAG 53536
184 laArgGlyTyrProProArgTrpProSerArgProProProSerCysTrp 200
|||||
53535 CTCGCGGCTGCGCTTCGCGCAGGTGGCTTCGCGGCGGCTTCCTGCTGG 53486
201 CysSerArgProThr 205
|||||
53485 TGCTCGGCGGCGCAGC 53471

```

seq_name: gb_hhg:AL607089

seq_documentation_block:

LOCUS AL607089 194215 bp DNA linear HTG 30-JAN-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-24819, *** SEQUENCING IN
PROGRESS ***, 2 ordered pieces.
ACCESSION AL607089

VERSION AL607089.15 GI:18476794
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE Brown,J.
AUTHORS Direct Submission
TITLE Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced g1:17402315.

COMMENT

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba24819
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 192819 bases at least Q40
Consensus quality: 193115 bases at least Q30
Consensus quality: 193255 bases at least Q20
Insert size: 194215; sum-of-ctdigs
Insert size: 193590; 7.2% error; agarose-fp
Quality coverage: 9.54x in Q20 bases; sum-of-ctdigs Quality
coverage: 9.62x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 16618: contig of 16618 bp in length
* 16619 16718: gap of 100 bp
* 16719 194215: contig of 17497 bp in length.
Location/Qualifiers

FEATURES

source

1. 194215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-24819"
/clone_1b="RP11-11.1"
1. 194215
/note="assembly_fragment:05205
clone_end:sp6
vector_side:left
clone_end:t7
vector_side:right"

misc_feature

BASE COUNT 55302 a 40938 c 40961 g 56913 t 101 others
ORIGIN

alignment_scores:
Quality: 1119.00 Length: 205
Ratio: 5.459 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-899-513-2 x AL607089/rev ..

Align seg 1/1 to reverse of: AL607089 from: 1 to: 194215

```

1 MetGlyAlaAlaPheValAlaSerLeuArgSerAsnLeuSerSerAlaThr 17
|||||

```

seq_name	gb	pr	AF080214
seq_documentation_block		1534 bp	mRNA linear
LOCUS	AF080214		
DEFINITION	Homo sapiens	protease-activated receptor 4	mRNA, complete cds.
ACCESSION	AF080214		
VERSION	AF080214.1	GI:3396080	
KEYWORDS	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Kahn, M.L., Hammes, S.R., Botka, C. and Coughlin, S.R.		
JOURNAL	Gene and locus structure and chromosomal localization of the		
REFERENCE	protease-activated receptor gene family		
AUTHORS	J. Biol. Chem. (1998) In press		
JOURNAL	2. (bases 1 to 1534)		
REFERENCE	Kahn, M.L., Zheng, Y.W., Huang, W., Bigornia, V., Zeng, D., Moff, S.,		
AUTHORS	Parsee, R.V., Tam, C. and Coughlin, S.R.		
JOURNAL	A dual thrombin receptor system for platelet activation		
REFERENCE	Nature (1998) In press		
JOURNAL	3. (bases 1 to 1534)		
REFERENCE			

[illegible]

```

158 yAlaProArGaLaLeuProAla..... 165
166 GCGTCGATCGCGTCTGCTGCTGATGACGGCGCCCGCGAGCGACAGGC 715
166 .....ProSerProGlyTTPArGaLaTrp..... 173
171 CTCCCACTGGCAACCGCGCTTCACCTGCGCTGCGCTGGCGCTGTTCC 765
174 .....ProLeuProAlaTr 178
176 TGCCCTGCTGCGCATGCTGCTGCTACGGGGCCACCTGCACACGCTG 815
178 pSerThraLaGlyAlaLeuArgLy..... 186
186 GCGGCGACGCGCGCGCTACGGCCAGCGGCTGAGCGACAGTGGT 865
187 ..TTPProProArGaTrpProSerArG...ProProSerCysTrpCys 201
866 GCTGCGCTGCGCG...TGCGCTTCTGCTGCGCGACAGCACTGCTGCTGC 912
name: gb_pat:AR108643

```

```

seq_documentation_block:
LOCUS AR108643 4895 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6111075.
ACCESSION AR108643
VERSION AR108643.1 GI:12824130
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4895)
AUTHORS Xu, W., Presnell, S. R., Yee, D. P. and Foster, D. C.
TITLE Protease-activated receptor PAR4 (ZCHEMR2)
JOURNAL Patent: US 6111075-A 1 29-AUG-2000;
FEATURES
source 1..4895
/organism="Unknown"
BASE COUNT 971 a 1573 c 1400 g 951 t
ORIGIN

```

```

alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

```

Alignment_block:

-09-899-513-2 x AR108643 ..

Align seg 1/1 to: AR108643 from: 1 to: 4895

```

40 ProAlaArGaPheLeuAlaValProAlaLaTrpAlaLeuAlaLeuGlyLe 56
401 CCCACACAG...CTGCTGCGCGCTGATGAGCGCTGCTGCTGCTGCTG 444
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaLeuPheLeuArgSerGly 73
445 GGGCTGCGCGCATGCGCTGCGCTGCGCTGCGCTGCGCGACGCGC 494
73 LyArgLeuGlyAlaAlaLeuLeuTyTrpLeuPheAsnLeuAlaLeuVal 89
495 CTCGCTGCGCTGCGCATGCGT...CTGATACACTGCGACTGCT 538
90 AspGluPhePheThrLeuThrLeuGlnLeuTrpLeuThrTyTrpLeu. 105
539 GACCTGCTGCTGCGCTGCGCGCTGCGCGCGCGCGCTGCGCTGCGCG 588
106 .....GlyLeuAlaArGaTrpProAlaLaTrpArGaTrpProGlyPro 118
589 TGCGCAGCGCTGCGCTGCGGAGGCGCGCTGCGCGCTGCGCGCGCGCG 638

```

```

119 ProThr.....ThrCysProPrometArGaTr 127
639 CACTCTATGTCATCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 688
127 gTrp.SerSerProArGSerSerAlaCysAlaAlaAlaLaserTyAla 143
689 CTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
144 ValProGlyProGlyArgLeu...ProAlaTrp..... 153
739 TGCGCGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 788
154 .....ProGlyAlaTyTrp 158
789 CCGCGCTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
158 yAlaProArGaLaLeuProAla..... 165
839 GCGTCGATCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888
166 .....ProSerProGlyTTPArGaLaTrp..... 173
889 CTCCCACTGGCAACCGCGCTTCACCTGCGCTGCGCTGCTGCTGCTGCT 938
174 .....ProLeuProAlaTr 178
939 TGCCCTGCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 988
178 pSerThraLaGlyAlaLeuArgLy..... 186
989 GCGGCGACGCGCGCGCTACGGCCAGCGGCTGAGCGTACCGCAGTGGT 1038
187 ..TTPProProArGaTrpProSerArG...ProProSerCysTrpCys 201
1039 GCTGCGCTGCGCG...TGCGCTTCTGCTGCGCGACAGCACTGCTGCTGC 1085
seq_name: gb_pr:AF055917
seq_documentation_block:
LOCUS AF055917 4895 bp mRNA linear PRI 07-JUL-1998
DEFINITION Homo sapiens protease-activated receptor 4 mRNA, complete cds.
ACCESSION AF055917
VERSION AF055917.1 GI:3293321
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4895)
AUTHORS Xu, W., Andersen, H., Whitmore, T. E., Presnell, S. R., Yee, D. P.,
Ching, A., Gilbert, T., Davie, E. W. and Foster, D. C.
TITLE Cloning and characterization of human protease-activated receptor 4
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6642-6646 (1998)
MEDLINE 96283984
REFERENCE 2 (bases 1 to 4895)
AUTHORS Xu, W., Andersen, H., Whitmore, T. E., Gilbert, T., Davie, E. W. and
Foster, D. C.
TITLE Direct Submmission
JOURNAL Submitted (26-MAR-1998) Biochemistry, University of Washington, J
wing, NE Pacific Street, Seattle, WA 98195, USA
FEATURES
source 1..4895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p12"
176..1333
/note="PAR4; member of family of G-protein coupled
protease activated receptors"
/product="protease-activated receptor 4"
/protein_id="AAC25689.1"
/db_xref="GI:3293322"

```


[illegible]

repeat_region	variation	/gene="F2RL3" /frequency="0.01" /replace="A" 2656. .2823
	repeat_region	/rpt_family="MTR1C" /rpt_type=dispersed 2757
	variation	/gene="F2RL3" /frequency="0.21" /replace="A" 2824. .3101
	variation	/rpt_family="AluIub" /rpt_type=dispersed 2852
repeat_region	variation	/gene="F2RL3" /frequency="0.01" /replace="A" 3021
	variation	/gene="F2RL3" /frequency="0.03" /replace="A" 3063
	variation	/gene="F2RL3" /frequency="0.02" /replace="A" 3108
	repeat_region	/gene="F2RL3" /frequency="0.01" /replace="C" 3154. .3470
variation	repeat_region	/rpt_family="AluIo" /rpt_type=dispersed 3282
	variation	/gene="F2RL3" /frequency="0.01" /replace="T" 3327
	variation	/gene="F2RL3" /frequency="0.12" /replace="T" 3481. .3788
	repeat_region	/rpt_family="AluSg" /rpt_type=dispersed 3564
variation	variation	/gene="F2RL3" /frequency="0.01" /replace="T" 3761
	variation	/gene="F2RL3" /frequency="0.02" /replace="T" 3794. .4085
	repeat_region	/rpt_family="MTR1C" /rpt_type=dispersed 4202
	variation	/gene="F2RL3" /frequency="0.01" /replace="G" 4587
variation	variation	/gene="F2RL3" /frequency="0.01" /replace="C" 4611
	variation	/gene="F2RL3" /frequency="0.21" /replace="T" 4759
	variation	/gene="F2RL3" /frequency="0.04" /replace="A" 4765
	variation	/gene="F2RL3" /frequency="0.07"

```
repeat_region
  /replace="T"
  /freq="4988"
  /rbp_family="L1MD1"
  /rpe_type="dispersed"
variation
  /gene="F2RL3"
  /frequency="0.35"
  /replace="T"
variation
  /gene="F2RL3"
  /frequency="0.22"
  /replace="C"
variation
  /gene="F2RL3"
  /frequency="0.03"
  /replace="A"
variation
  /gene="F2RL3"
  /freq="6236"
```

```
alignment_scores:
    Quality: 177.50      Length: 235
    Ratio:   1.659       Gaps:    13
Percent Similarity: 45.538 Percent Identity: 30.638
```

```
alignment_block:
US-09-899-513-2 x AF384819 .
```

Align seg 1/1 to: AF384819 from: 1 to: 11828

[illegible]

174ProleuproAlaTr 178
174
2244 TGCCCTGCTGCTGCTGCTACGGGGACCCCTGACACGCTG 2293
178 pserThraAlaGlyAlaAArgly..... 186
178
2294 GCGGCACAGCGCGGCTACGGCCAGCGCTGAGCGTACCGCAGTGT 2343
187 ..TTPProProArgrTTPProSerArg...ProProSerCysTTPCys 201
187
2344 GCTGGCCCTCCGCCG...TGCCCTTCTGTGTGCTCCAGCAACCTGCTGCTGC 2390
seq_name: gb_htg:AC008737
seq_documentation_block:
LOCUS AC008737 248680 bp DNA linear HTG 26-JUL-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-253869, *** SEQUENCING IN
PROGRESS ***, 2 ordered pieces.
ACCESSION AC008737 GI:15021998
VERSION AC008737.9
WORDS HTG: HTGS_PHASE2; HTGS_ACTIVEFIN.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 248680)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 248680)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 26, 2001 this sequence version replaced gi:13752640.
* NOTE: This is a 'working draft' sequence.
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and -
* the accession number will be preserved.
* 1 155093: contig of 155093 bp in length
* 155094 155193: gap of unknown length
* 155194 248680: contig of 93487 bp in length.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 804726
Center clone name: CTD-253869

Summary Statistics
Consensus quality: 247465 bases at least Q40
Consensus quality: 247949 bases at least Q30
Consensus quality: 247983 bases at least Q20
Estimated insert size: 248980; agarose-fp estimation
Estimated insert size: 248035; sum-of-contigs
estimation
Quality coverage: 11.93 in Q20 bases; agarose-fp
estimation
Quality coverage: 11.97 in Q20 bases; sum-of-contigs
estimation.
Location/Qualifiers
1. 248680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-253869"

BASE COUNT 63562 a 62153 c 62761 g 60104 t 100 others
ORIGIN
alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent similarity: 45.532 Percent identity: 30.638
alignment_block:
US-09-899-513-2 x AC008737/rev ..
Align seg 1/1 to reverse of: AC008737 from: 1 to: 248680
40 ProAlaArpHeIleValIValProAlaAlaTyrAlaAlaLeuAlaGlyLe 56
|||||
42859 CCCACAGG.....CTGCTGCCGCCCTCTATGGCGTGTCTGTGTGT 42816
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGly 73
:|||||
42815 GGGCGTCCGCCCATGTGGCTGTGGCTGTGGCTGTGGCCACGACGAC 42766
73 lYArgLeuGlyAlaAlaLeuLeuLeuTyrLeuPheAsnLeuAlaVal 89
|||||
42765 CTCGCTCCCTCTCCACATGCTG.....CTGATGAACCTCGCGCTCT 42722
90 AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuThrTyrLeu 105
|||
42721 GACCTCTGTGCGCCCTGCGCGTGGCCCGCGGATCGCTACACCTCG 42672
106GlyLeuAlaArgrProProAlaThrArgProGlyPro 118
|||
42671 TGCCAGCGCTGCGCTTCGGGAGCGCGCTGCGCTGCGAGGCGG 42622
119 ProThr.....ThyProProMetArgAr 127
42621 CACTTATAGTCACATGTATGGCTCAGTCTGCTGCGCGCTCAGC 42572
127 gTTP_SerSerProArgrSerSerAlaCysAlaAlaAlaSerTyrAla 143
|||
42571 CTGATTCCTACCTGCGCTGCGCTGTCACCGCGCGCGCGCTGCG 42522
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
|||||
42521 TGCGCGCGCTGCGCTTGGACTCTGCATGCTGCTGCTCATGCGG 42472
154ProGlyAlaTyrG 158
42471 CGGCCCTGACCTGCCCTGACACTGCAGCGGACAGACCTTCGCGTGGC 42422
158 yAlaProArgrAlaLeuProAla..... 165
|||||
42421 CGCTCGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42372
166ProSerProGlyTTPArgAlaTrp..... 173
42371 CTCCCATGTGCAACCGCGCTTACCTGCTGCTGCTGCTGCTGCT 42322
174ProleuproAlaTr 178
42321 TGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 42272
178 pserThraAlaGlyAlaAArgly..... 186
178
42271 GCGGCACAGCGCGGCTACGGCCAGCGCTGAGCGTACCGCAGTGT 42222
187 ..TTPProProArgrTTPProSerArg...ProProSerCysTTPCys 201
187
42221 GCTGGCCCTCCGCCG...TGCCCTTCTGTGTGCTCCAGCAACCTGCTGCTGC 42175
seq_name: gb_dr:HSU91939
seq_documentation_block:

LOCUS	HSU91939	1400 bp	DNA	linear	PRI 22-MAR-1997
DEFINITION	Human putative G protein-coupled receptor (GPR25) gene, complete cds.				
ACCESSION	U91939				
VERSION	U91939.1	GI:1905877			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1400) Jung,B.P., Nguyen,T., Kolakowski,L.F. Jr., Lynch,K.R., Heng,H.H., George,S.R. and O'bowd,B.F.				
TITLE	Discovery of a novel human G protein-coupled receptor gene (GPR25) located on chromosome 1				
JOURNAL	Biochem. Biophys. Res. Commun. 230 (1), 69-72 (1997)				
MEDLINE	2 (bases 1 to 1400)				
REFERENCE	Jung,B.P., Nguyen,T., Kolakowski,L.F. Jr., Lynch,K.R., Heng,H.H., George,S.R. and O'bowd,B.F.				
AUTHORS	Direct Submission				
TITLE	Submitted (26-FEB-1997) Pharmacology, University of Toronto, 8 Taddle Creek Rd., Toronto, ON M5S 1A8, Canada				
JOURNAL	location/Qualifiers				
FEATURES	1..1400				
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="1q32.1" 80..1162 /gene="GPR25" 80..1162 /gene="GPR25" /codon_start=1 /product="putative G protein-coupled receptor" /protein_id="AAB50155.1" /cd_xref="GI:1905878" /translation="MAPLEWSPSPGSAWPDYSGLDGLELELCOPAGDLPGYVYIPRLYLAAVAGLGNAFVVMVLLAGRGPRRLVDTEVLHLAAADLGFLVTLPLMAAARAR PWPGDGLCKSTFALAGTRSGALLAGMSVDRYLVAKLEERPLRTPCVAASCC GMAVVALLAGLPSLVYRGLOLPDGGOSGCEESHAFQGLSILLITLTPVLYVTL PCYCRISRRLRPRPHVGRARANSRIITPAISTVVGSLPFSARAYHLARLALP PCPLTLARWGLTITATCALAYNSCANPLITLLDRSFARALDAGCGTGLARISIS ASSISRDSSVFERKAADANVASAW"				
BASE COUNT	154 a	524 c	474 g	248 t	
ORIGIN					
alignment_scores:					
Quality:	171.00	Length:		212	
Ratio:	1.583	Gaps:		12	
Percent Similarity:	50.943	Percent Identity:		32.547	
Alignment block:	US-09-899-513-2 x HSDU91939 ..				
Align seg 1/1	to: HSU91939	from: 1	to: 1400		
24	SerSerValGlyAspLeuGlyValGly..	C	34	
118	AGCGCCCTGGACTACTCGGGGTGGACGCGCTCGACGAGACGTGGACCTGT	167			
34	ySerLeuTrpAspAspProAlaIarGPhelIleValValProAlaIaIarY	50			
168	GTCCGCGCGCGGACCTCGCCCTACGCGTACGTACTACATCCCGCGGCTCAG	217			
51	AlaLeuAlaLeuGlyLeuGlyLeuPProAlaAsnValAlaAlaLeuAlaLeu	67			
218	CTGGCGGCGCTCGCCCTGGCGCTGCTGGCGCAC.....GCCTTTGTGG	261			
67	ThpeIleArgSerGlyArgLeuGlyGlnAlaLeuLeu..LeuTyrL	83			
262	GTGGCTGCTGGCGCGCGCGCGCGCGCGCTGTGTGGATACCTTGC	311			

```

83 euphasinLeuValaspGluPhePheGluLeuGlnLeu 99
:::||||| ::||| ||:||||||| ||
312 TGTGTCACTGGCGGAGCTGACCTGGGCTGTGTCTCAGCTGCCGCTG 361
::: :::: ||||| |||
362 TGGGCGCG.....GGGGGCTAGGCGGCGCTG..... 390
::: ::||| |||
116 OGlyProPThrThCysProBromEtkArgArgTyrSerSerProArgS 133
::: |||||
391 .....GCCCTGGCGGATGG.....CCCTGC 412
::: |||||
133 eSerAlaCysAlaAlaAlaAlaSerTyrAlaValProGly..... 146
::: ::|||:|||||
413 AAGCTCAGACGCTGGCGGCTGGCGGAGCAGCGCTGGCGGCGGCTGCT 462
147 .....ProGlyArgLeuProAlaThrPr 154
||||| |||
463 GCTGGCGGCGATGAGCGTGGACCGCTACCTGGCGCGGTGAAGCTGCTG 512
154 OGly..AlaTyrGlyAlaProArgAlaAlaProAlaPro..... 166
||| :::: ||||| ||| |||
513 AAGCGAGGCGACTGGCGGACCCGCGCTGGCGCGCTGCTGCTGCGGCG 562
167 .SerProGlyTyrPArgAlaTyrPro...LeuProAlaTyrSerThrAla 182
||| |||||:||||| || |||||:|||||
563 GCTGTGGCGGTGGCGGCTGCTGGCGGCGCTGCCCTCCTGTCTACAGGG 612
182 Ty.....GlnAlaArgGlyTyrProProPro 190
||| ::| |||
613 GTTGACAGCCCTGCCTGGGGGCGAGGACGACGAGTGGCGGAGACCCCT 662
191 ArgTyrProSerArgProProSerCysTyrCys 201
||||||| ::||| |||
663 CCCACGCGCTTCAGGGCGCTCAGCTTGCCTGCTGC 695

seq_name: gb_pr:HUMCHRA

seq_documentation_block:
LOCUS HUMCHRA 1863 bp mRNA linear PRI 01-NOV-1994
DEFINITION Human Chromogranin A mRNA, complete cds.
ACCESSION J03483
VERSION J03483.1 GI:180526
KEYWORDS Chromogranin A.
SOURCE Human pheochromocytoma, cDNA to mRNA, (library of A. Lamoureaux and
J. Mallet), clone hc9H/42.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1863)
Konecki,D.S., Benedum,U.M., Gerdts,H.H. and Hultner,W.B.
The primary structure of human chromogranin A and pancreastatin
J. Biol. Chem. 262 (35), 17026-17030 (1987)
88059106
COMMENT Printed copy of sequence for [1] kindly provided by Konecki,D.S.,
12/15/87.

FEATURES
Source
1..1863
Location/Qualifiers
1..1863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14q32"
<1..1863
/gene="CHGA"
/product="Chromogranin A mRNA"
1..1863
/gene="CHGA"
/region="CHGA"
83..1456
/region="CHGA"
/note="Chromogranin A precursor"
/codon_start=1
/protein_id="AAA52017.1"
/db_xref="GI:180527"

```


Consensus quality: 156049 bases at least Q20
Insert size: 195000; agarose-fp
Insert size: 157491; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a working draft sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1373:	contlg of 1373	bp	in	length
*	1374	1473:	gap of	100	bp	in
*	1474	3371:	contlg of 1898	bp	in	length
*	3372	3471:	gap of	100	bp	in
*	3472	5991:	contlg of 2520	bp	in	length
*	5992	6091:	gap of	100	bp	in
*	6092	8001:	contlg of 1910	bp	in	length
*	8002	8101:	gap of	100	bp	in
*	8102	10452:	contlg of 2351	bp	in	length
*	10453	10552:	gap of	100	bp	in
*	10553	13499:	contlg of 2947	bp	in	length
*	13500	13559:	gap of	100	bp	in
*	13600	17300:	contlg of 3701	bp	in	length
*	17301	17400:	gap of	100	bp	in
*	17401	23167:	contlg of 5767	bp	in	length
*	23168	23267:	gap of	100	bp	in
*	23268	27334:	contlg of 4127	bp	in	length
*	27395	27494:	gap of	100	bp	in
*	27495	31941:	contlg of 4447	bp	in	length
*	31942	32041:	gap of	100	bp	in
*	32042	36867:	contlg of 4826	bp	in	length
*	36868	36967:	gap of	100	bp	in
*	36968	43237:	contlg of 6270	bp	in	length
*	43238	43337:	gap of	100	bp	in
*	43338	52795:	contlg of 9458	bp	in	length
*	52796	52895:	gap of	100	bp	in
*	52896	61446:	contlg of 8551	bp	in	length
*	61447	61546:	gap of	100	bp	in
*	61547	68822:	contlg of 7276	bp	in	length
*	68823	68922:	gap of	100	bp	in
*	68923	79907:	contlg of 10985	bp	in	length
*	79908	80007:	gap of	100	bp	in
*	80008	91003:	contlg of 10936	bp	in	length
*	91004	91103:	gap of	100	bp	in
*	91104	109448:	contlg of 18325	bp	in	length
*	109429	109528:	gap of	100	bp	in
*	109529	135277:	contlg of 25749	bp	in	length
*	135278	135377:	gap of	100	bp	in
*	135378	159391:	contlg of 24014	bp	in	length

FEATURES
SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone="RP11-67212"
/clone_lib="RP11 Human Male BAC"
1..1373
/misc_feature
/note="assembly-fragment"
1474..3371
/misc_feature
/note="assembly-fragment"
3472..5591
/misc_feature
/note="assembly-fragment"
6092..8001
/misc_feature
/note="assembly-fragment"
8102..10452
/misc_feature
/note="assembly-fragment"
10553..13459
/misc_feature
/note="assembly-fragment"

```

```
misc_feature 13600. 17300 /note="assembly_fragment"  
misc_feature 17401. 23167 /note="assembly_fragment"  
misc_feature 23268. 27394 /note="assembly_fragment"  
clone_end:T7 vector_side:right"  
misc_feature 27495. 31941 /note="assembly_fragment"  
misc_feature 32042. 36867 /note="assembly_fragment"  
misc_feature 36968. 43237 /note="assembly_fragment"  
misc_feature 43338. 52795 /note="assembly_fragment"  
misc_feature 52896. 61446 /note="assembly_fragment"  
misc_feature 61547. 68822 /note="assembly_fragment"  
misc_feature 68923. 79907 /note="assembly_fragment"  
misc_feature 80008. 91003 /note="assembly_fragment"  
misc_feature 91104. 109428 /note="assembly_fragment"  
misc_feature 109529. 135277 /note="assembly_fragment"  
misc_feature 135378. 159391 /note="assembly_fragment"
```

	vector_side:right"									
BASE COUNT	43727	a	35737	c	36282	g	41744	t	1901	others
ORIGIN										

```
alignment_scores:      .
    quality:          167.50   Length:       200
    Ratio:            1.565     Gaps:         10
Percent Similarity:    53.500   Percent Identity: 34.500
```

```
alignment_block:
US-09-899-513-2 x AC027113 .
```

Align seg 1/1 to: AC027113 from: 1 to: 159391

```

24 ServerValGlyAspleuGlyValGly.Gly.....C 34
    |||:::||||| |||
132146 AGCGCCCTGGGACTACTCGGGCTTGCAGCGCCTGAGAGACTGGAGCTGT 132195

```

34 ySserLeuTrpaspAspProAlaarghelleValValProAlaalaTyr 50
 || ||| ||| ::::: ::||| |||
 132196 GTCCGGCCGGGACCTGCCCTACGGCTACGTTACATCCCGCGCTTAC 132245

51 AlaleualaleuGl vLeuGl vLeuProAlaasvAlaAlaleuAlame 67

1322246 CTGGCGCCCTTCGCCGTGGGCTGCTGGGCAAC....GCCCTTGCTGT 1322289

67 tpheileargserglyargleuglyglnalaleu...LeuTyrl 83

132290 GTGGCTGTCGCGCCGCGGGCGGCCGCGCTGTGGATACCTTCG 132339

83 eupheasinleuAlaleuValaspcluphenehThreuthreuglnleu 99

132340 TGGTGCACCTGGCGGCAGCTGACCTGGGCTTCGTGCTACAGCTGCCGCTG 132389

[illegible]

132390 TGGGCGCGCTAGGCGGCGCTGGCGTTCGGCG 132432

```

116 oglyProThrThrcysProPromeIargTTPserSerProArgS 133
      |||      ::      ::|||      ::

```

133 erSerAlaCysAlaAlaAlaSerTyraAlaValProGlyProGlyArg 149
132480 GCGGGCCCTGCTGCTGCGGGCAGTACGCGTACCGCTGCGCTG 132529
150 LeuProAlaTrpProGly..AlaTyrcGlyAlaProAlaLeu..... 163
132530 GGTGACCTCTCTGAGGCGAGGCCACTGCGCACCCCGCTGCGCGCTG 132579
164ProAlaProSerProGlyTrpArgAlaTrpPro...LeuProAl 177
132580 CCGCTGCTCTGCGGCTGCTGCGGCTGCGCTGCGCGCGCTGCTCC 132629
177 aTrpSerThaAlaGly.....GlnAlaArg 186
132630 CTGCTACCGGGGCTTGACAGCCCTGCTGCGGCGGCGAGACAGCAGTG 132679
186 LyTrpProProAlaTrpProSerArpProProSerCysTrpCys 201
132680 CGGCGAGAGCCCTCCACGCTTCAGGCGCTCAGCTGCTGCTGC 132726
_name: gb_pr:AC099756
seq_documentation block:
LOCUS AC099756 167302 bp DNA linear PRI 25-JAN-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-180A14, complete sequence.
ACCESSION AC099756 AL356097
VERSION AC099756.2 GI:18369993
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167302)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
Direct Submission
Unpublished
2 (bases 1 to 167302)
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
Direct Submission
Submitted (20-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 167302)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
Direct Submission
Submitted (25-JAN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Jan 25, 2002 this sequence version replaced gi:17017615.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-180A14 (sc0650)
----- Summary Statistics
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 167134 bases at least Q40
Consensus quality: 167259 bases at least Q30
Consensus quality: 167292 bases at least Q20
Insert size: 167302; sum-of-contigs
Quality coverage: 12.3x in Q20 bases; sum-of-contigs

Overlapping sequences:
5': RP11-93N17 AL450104, 45836-bp overlap
3': RP11-168016 AL358473, 2000-bp overlap

----- Sequence Quality Assessment: -----
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation: -----

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI		BglII		HindIII	
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	8800	6289	6193	5506	5410
6	<800	2067	2160	6382	6417
5256	5421	10227	10337	512	<800
3894	3906	3493	3469	449	<800
7708	7866	8686	8888	5504	5410
10361	10398	6857	6991	669	<800
3486	3515	1470	1453	9353	9325
8907	8800	6157	5791	2207	2275
4188	4148	2147	2160	198	<800
6893	6917	1765	1776	5094	5021
77	<800	571	<800	6008	6010
5468	5421	8095	8055	3067	3100
10689	10398	3267	3285	1366	1372
7761	7866	3082	3087	3297	3359
4544	4563	1042	1000	1641	1647
854	858	1130	1117	77	<800
5389	5421	2350	2427	933	938
5310	5421	4889	4937	3789	3753
1232	1233	3659	3616	7996	8013

145	<800	43	<800	915	938
130	<800	2429	2427	12536	12408
1052	1059	12011	12051	2246	2275
8669	8800	12123	12051	8092	8013
945	946	9551	9501	1220	1209
1539	1506	2525	2561	4710	5021
5862	5822	6231	6193	14012	14309
2871	2990	3365	3469	13389	13285
8216	8211	10028	9966	295	<800
9331	8800	1222	1204	587	<800
1197	1233	3361	3285	5028	5021
1315	1233	288	<800	6162	6010
3266	3215	4825	4798	344	<800
2453	2472	2188	2160	296	<800
7302	7228	5912	6193	1009	938
16816	17015	975	1000	7219	7087
214	<800	5851	5791	5930	6010
1818	1825	6279	6193	2266	2275
2121	2102	9551	9501	74	<800

FEATURES
Source Location/Qualifiers
1..167302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-180A14"
/clone_lib="RPC1 human BAC library 11"

BASE COUNT 43765 a 37479 c 38785 g 47273 t
ORIGIN

alignment_scores:

Quality: 167.50 Length: 200
Ratio: 1.565 Gaps: 10
Percent Similarity: 53.500 Percent Identity: 34.500

Alignment_block:

us-09-899-513-2 x AC099756 ..

Align seg 1/1 to: AC099756 from: 1 to: 167302

```
24 SerSerValGlyAspIleuGlyValGly.....C 34
|||||:|||||:|||||:|||||:|||||:
55197 AGCGCCTGAGACTACTCGGGGTGGACGCGCTGAGAGAGAGTGAAGCTGT 55246
34 ySerIleuTrpAspProAlaArgPheIleValValProAlaIleuTrp 50
|||||:|||||:|||||:|||||:|||||:
55247 GTCCGGCGGGGACCTCCCTACGCTACCTACCTACCTACCTACCTAC 55296
51 AlaIleuAlaIleuGlyLeuGlyLeuProAlaAsnValAlaAlaIleuAla 67
|||||:|||||:|||||:|||||:|||||:
55297 CTGGCGGCGCTTGGCGCTGGCGCTGGCGGAC.....GCCTTGTGTGT 55340
67 rPheIleArgSerGlyArgLeuGlyGlnAlaIleuLeu...LeuTyrL 83
|||||:|||||:|||||:|||||:|||||:
55341 GTGGCTGTGGCGGGCGGGCGGGCGGGCGGGCGGGCGGTGATACCTTGT 55390
83 eUpheAsnIleuAlaIleuValAspGlnPhePheIleuTrpLeuGlnIleu 99
|||||:|||||:|||||:|||||:|||||:
55391 TCGTCACCTGCGGCGGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 55440
100 TrpLeuThrTyrTyrLeuGlyLeuAlaArgArgProAlaThrArgPr 116
|||||:|||||:|||||:|||||:|||||:
55441 TGGCGCGCGGCG.....GGCGGCGTACGGCGCGCGCTTGTGGCG 55483
116 oGlyProProThrThrCysProPrometArgArgTrpSerSerProArgS 133
|||||:|||||:|||||:|||||:|||||:
55484 ATGG...CCTCGCAAGCTCAGCAGCTTGGCGTGGCGGCGGCGGCTGC 55530
133 eSerAlaCysAlaAlaAlaAlaAlaSerTyrAlaValProGlyArg 149
|||||:|||||:|||||:|||||:|||||:
55531 GCGGGCGCGCTGCTGCTGGCGGGCATGACCGTGGACCGCTACCTGCGCT 55580
150 LeuProAlaTrpProGly...AlaTyrGlyAlaProArgAlaLeu..... 163
|||||:|||||:|||||:|||||:|||||:
55581 GGTGAAGCTGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGG 55630
164 .....ProAlaProSerProGlyTrpArgAlaTrpPro...LeuProAl 177
|||||:|||||:|||||:|||||:|||||:
55631 CCTCGTGTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 55680
177 rTrpSerThrAlaGly.....GlnAlaArg 186
|||||:|||||:|||||:|||||:|||||:
55681 CTGGCTTACCGGGGTTGCGACCGCTGCGGGGCGGCGGCGGCGGCGGCT 55730
186 lYTrpProProArgTrpProSerArgProProSerCysTrpCys 201
:::|||||:|||||:|||||:|||||:
55731 CGGCGAGAGCGCTCCACGCGCTCCAGGCGCTCAGCTTGTCTGTCTGC 55777

seq_name: gb_pr:HUMCHRAA
seq_documentation_block:
LOCUS HUMCHRAA 1811 bp mRNA linear PRI 01-NOV-1994
DEFINITION Human chromogranin A mRNA, complete cds.
ACCESSION J03915
VERSION J03915.1 GI:180528
KEYWORDS chromogranin A; parathyroid secretory protein 1.
SOURCE Human pheochromocytoma, cDNA to mRNA, clone pHCGA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1811)
Helman,L.J., Ahn,T.G., Levine,M.A., Allison,A., Cohen,P.S.,
Cooper,M.J., Cohn,D.V. and Israel,M.A.
Molecular cloning and primary structure of human chromogranin A
```

(secretory protein 1) cDNA
JOURNAL J. Biol. Chem. 263 (23), 11559-11563 (1988)
MEDLINE 88298816
COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by L.J. Hellman, 31-MAY-1988.

FEATURES
source
Location/Qualifiers
1..1811
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14q32"
1..1811
/gene="CHGA"
43..1416
/gene="CHGA"
/note="chromogranin A"
/codon_start=1
/protein_id="AA52018.1"
/db_xref="GI:180529"
/translation="MSAAVLLALLCAGVATLPVNSPMNKGDTEVMKIVEITYDTL
SKSPMPVSKCEFTLRGDERITLIRHNLKELDLALOGKKERAHOKKSGPED
ELSEVLENSQALIKQAVPESSKDYMKRDSRAEKSGATGARGAOLPEPMOE
SKAEGNKQAPGEEEEEERATITHPASIPSKYRPOAKGSEGLSQGLVREKGLS
VPPGQAKREEEEEEERAEAEAVPEEGPTVVPNPISLGIKIRGESSEALA
VQAGKPGAEADPEGKGEQHSQKDEEEMAVVQGLFRGKSGLEQEEERLSK
EMEDNRMKMDQAKELTAERKRLGQEEEDNRDSMRSLSPFARGYGRGPGQLRR
GSRPNSMEDSLBAGLPLQVRYGYPQEKKEEGSANRRPEQLESLSAIIEALEKVAHQ
LOCALR"

BASE COUNT 432 a 509 c 599 g 271 t
ORIGIN Chromosome 14.

alignment_scores:
Quality: 164.50 Length: 219
Ratio: 1.750 Gaps: 12
Percent Similarity: 42.922 Percent Identity: 30.594

alignment_block:

US-09-899-513-2 x HUMCHRAA/rev ..

Align seg 1/1 to reverse of: HUMCHRAA from: 1 to: 1811

```

19 ArgSerGluMetAsnSerSerValGlyAspLeuGly..... 30
|||||: : : : : : : : : : : : : : : : : : : : :
1587 AGCAGTAGAGACAAAGCGCGCTAGGCGCATAGGTGGCTGGGCTTGG 1538
31 ....ValGlyGlyGly.....SerLeuTPA 38
|||||: : : : : : : : : : : : : : : : : : : : :
1537 GCAGGCTGGAGGCTGCTCCCTACCGGAGGAGCGGCGGCATCTGGC 1488
38 spAspProAlaArgPheIleValProAlaAlaTyrAlaLeuAlaLeu 54
|||||: : : : : : : : : : : : : : : : : : : : :
1487 CAGGACCTGCCAAGGAGGACAGACAGAGGCGCACAGGCT...GCCCTG 1441
55 GlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgS 71
||| : : : : : : : : : : : : : : : : : : : : :
1440 GGGCCAGCGCGTGCAGCGGTGTCTCAGCCCG.CCCGTAAGTGCCTGCAGC 1392
71 rGlyGlyArgLeuGlyAlaLeuLeuLeuTyrLeuPheAsnLeuAlaL 88
:||||| : : : : : : : : : : : : : : : : : : : :
1391 TGGTGGGCGCACCTTCTCCAGCT..... 1370
88 euValAspGluPhePheThrLeuThrLeuGlnLeuTyrLeuThr..... 102
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1369 .....CTGCCTCATGTGCTCACAAGCTCT 1346
103 .....TyrTyrLeuGlyLeuAlaArgArgProProAlaTh 114
||||| ||| ||| ||| ||| ||| ||| ||| :
1345 CCAGCTCTGCTCTGCTGCTGGGTTTGGCGTGCCTCTCTCTTTC 1296
114 rArgPro.....GlyProProThrThrC 122
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1395 TTCTCTGGGGGTAGCTCGGAGCTGAGGGGCAAGGCGCGCTCAAGGCT 1246

```

```

122 ysProProMetArgArgTyrPheSerSerProArgSerSerAlaCysAlaAla 138
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1245 GTCTT.....CCAGAGAGTGGCGCTCGAGGCTC 1217
139 AlaAlaSerTyrAla.....ValProGlyProGlyArg 149
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1216 GCCCGAGCTGCGGCCAGAGCCCTGAGAGCCGCGGCGCGGAG 1167
149 GluProAlaTyrPhe.....ProGlyAlaTyrGlyAlaProArgAlaLeuP 164
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1166 GAGAGCTTCATGAGAACGCTCCGGTGTCTCTCTCTCTCTCTCTCTCTCTC 1117
164 roAlaProSerProGlyTyrPargAlaTyrProLeu...ProAlaTyrSer 179
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1116 CAGCGCTTCTCAGCCGCTGAGCTCTTGGCAGCGTGCATCTTGTCTCC 1067
180 ThrAlaGlyAlaAlaArgGlyTyr...ProProProArgTyrProSerArg 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1066 AGC.....GATTGGTGTCTCTCCACTCTCTGAGAGCCGC 1032
195 gProPro 197
|||||
1031 TCCTCT 1025
seq_name: gb_pr:BC012755

```

seq_documentation_block:
LOCUS BC012755 1432 bp mRNA linear PRI 20-AUG-2001
DEFINITION Homo sapiens, clone MGC:16126 IMAGE:3627785, mRNA, complete cds.
ACCESSION BC012755
VERSION BC012755.1 GI:15215328
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadn@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAL Plate: 24 Row: m Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source
Location/Qualifiers
1..1432
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:16126 IMAGE:3627785"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH-MGC_19"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
23..1009
/codon_start=1

alignment_block:
us-09-899-513-2 x AAX90981 ..
Align seg 1/1 to: AAX90981 from: 1 to: 1534

```

40 ProAlaArgPheIleValValProAlaIAlaIAlaLeuAlaLeuGlyLe 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 CCCACGAGG.....CTGCTGCCCGCCCTCTATGCGCTGCGCTGCGGT 271
56 uGlyLeuProAlaAsnValAlaIAlaLeuAlaMetPheIleArgSerGly 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
272 GGGGCTGCTCCGCAATGAGGCTGGCGCTGCTGCTGCTGCCACGAGGAC 321
73 LysArgLeuGlyAlaIAlaLeuLeuLeuThrLeuPheAsnLeuAlaLeuVal 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 CTGGGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGCAGCATGCT 365
90 AspGluPhePheThrLeuThrLeuThrLeuThrLeuThrLeuLeuLeu 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 GACCTCTGCTGGCGCTGCGCTGCCCGCGGATGCGCTACACCATGCGG 415
106 .....GlyLeuAlaArgArgProProAlaIAlaIAlaIAlaIAla 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 TGCGCCAGCGCTGGCCCTTCCGGGAGGCCGCTGCCCTGCCACGAGCG 465
119 ProThr.....ThrcysProPheMetArgArg 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466 CACCTCTATGTCACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
127 GTrp.SerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTrpAla 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 CTGATGCTGCTACCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 565
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
566 TGCGCGCGCGCTGGCCCTTGGACCTGCTGCTGCTGCTGCTGCTGCTG 615
154 .....ProGlyAlaTrpG1 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
616 CCGCGCTGGCACTGCCCTGACACTGACGCGGACAGACCTTCCGCTGCG 665
158 YAlaProArgAlaLeuProAla..... 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
666 CGCTCCGATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 715
166 .....ProSerProGlyTrpArgAlaTrp..... 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 CTCCACACTGCAACCGGCTTACCTGCTGCTGCTGCTGCTGCTGCTG 765
174 .....ProLeuProAlaTr 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
766 TGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 815
178 PserThrAlaGlyAlaAlaArgGly..... 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
816 GCGCGCCACGCGCGCGCTGACGCGCGCTGAGGCTGACCGCAGTGT 865
187 ..TrpProProArgTrpProSerArg...ProProSerCysTrpCys 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
866 GCTGCGCTCCGCGG...TGCGCTTCTTGTGCTGCGCCAGCACTGCTGCTG 912

```

seq_name: /SIDSI/gcgdata/09/seq/geneseqn-emb1/NA2000.DAT:AAF21431
seq_documentation_block:
ID AAF21431 standard; DNA: 1534 BP.

AC AAF21431:
XX
XX
XX 14-MAR-2001 (first entry)
XX
XX Human low adenosine antisense oligonucleotide related sequence #2998.
DE
XX

KW Low adenosine antisense oligonucleotide: phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiallergic; hypotensive; cytoskeletal;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
XX WO200062736-A2.
XX
XX 26-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
XX
XX Nyce JW;
XX
XX WPI; 2000-679539/66.
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX disclosure; Page 1415; 1592pp; English.
XX
XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (I) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'universal' or alternative base-
XX (I) can have respiratory, bronchodilator, antihypertensive, analgesic,
XX immunosuppressive, antiallergic, hypotensive and cytoskeletal activities.
XX The antisense oligonucleotides and (I) can be used to down-regulate the
XX expression and/or activity of target polypeptides associated with
XX lung/respiratory disorders and malignancies, such as stimulating and
XX activating peptide factors and transmitters, transcription factors,
XX immunoglobulins and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central
XX nervous system (CNS) and peripheral nervous and non-nervous system
XX receptors, CNS and peripheral nervous and non-nervous system peptide
XX transmitters, defensins, growth factors, vasodilative peptides and
XX receptors, binding proteins and malignancy associated proteins. The
XX antisense oligonucleotides may be used in this way to treat disorders
XX including respiratory obstruction (especially pulmonary obstruction
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
XX and/or surfactant hypoproduction which are associated with a disease or
XX condition selected from pulmonary vasoconstriction, inflammation,
XX allergies, asthma, impeded respiration, respiratory distress syndrome
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,
XX and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
XX fragments and antisense oligonucleotides used in the exemplification of
XX the present invention.
XX
XX Sequence 1534 BP; 235 A; 526 C; 479 G; 293 T; 1 other;

alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638
alignment_block:

US-09-899-513-2 x AAF21431 ..
Align seg 1/1 to: AAF21431 from: 1 to: 1534

```

40 ProAlaArgPheIleValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
228 CCCACCAGG.....CTGGTCCCGCCCTCTATGGGCTGCTGCTG 271
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGlyG 73
272 GGGCTGCTGGCCCAATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 321
73 LyArgLeuGlyAlaAlaLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
322 CTCGGCTCCCTCCACCATGCTG.....CTCATGACCTCCGCACTGCT 365
90 AspGluPhePhePheLeuThrLeuThrLeuThrLeuThrLeuThrLeu 105
366 GACCTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 415
106 .....GlyLeuAlaArgProProAlaThrArgProGlyPro 118
416 TGGCCAGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 465
119 ProThr.....ThrCysProProMetArgArg 127
466 CACTATAGTGCATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
127 gTTPSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
516 CTGGATCCCTTACCTGGCCCTGGCTGCACCCGGCCGGCCGGCCCTGC 565
144 ValProGlyProGlyArgLeu..ProAlaTTP..... 153
566 TGGCCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 615
154 .....ProGlyAlaTyrG1 158
616 CCGCCCTGGCACTGGCCCTGCACATGCACGCGGACACCTTCGGGCTGG 665
158 ValAlaProAlaAlaLeuProAla..... 165
666 CGCTCCGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715
166 .....ProSerProGlyTTPArgAlaTTP..... 173
716 CTCCTCCATGGCAACCGGCTTCACCTGCTGGCTGGCTGGCTGGCTGG 765
174 .....ProLeuProAlaTTP 178
766 TGGCCCTGCTGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 815
178 pSerThrAlaGlyAlaAlaArgGly..... 186
816 GCGGCGACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 865
187 ..TTPProProArgTTPProSerArg..ProProSerCysTTPCys 201
866 GCTGGCTTCCTGGCG...TGGCTTCTGCTGGCCAGCAACCTGCTGCTGC 912
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA35309
seq_documentation_block:
ID AAA35309 standard; DNA; 1534 BP.
XX
XX AAA35309;
XX
XX 28-JUL-2000 (first entry)
XX
XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:183.
XX
XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
KW

```

```

KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytotatic; analgesic; impaired airway;
KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukemia; lymphoma; carcinoma; metastasis; ss.
XX
XX Homo sapiens.
XX
XX WO200009525-A2.
XX
XX 24-FEB-2000.
XX
XX 03-AUG-1999; 99WO-US17712.
XX
XX 03-AUG-1998; 98US-0095212.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX NYCE JW;
XX
XX WPI; 2000-205971/18.
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers.
XX
XX Disclosure; Page 1330; 1343pp; English.
XX
XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytotatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
XX carcinomas, and cancers which may metastasize to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing the
XX bronchoconstriction and inflammation. AAA3313 to AAA35312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680
XX (AAA3233 to AAA3392) are specifically claimed ONs from the present
XX invention. N.B. Sequences given in the disclosure of the present
XX invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.
XX
XX Sequence 1534 BP; 235 A; 526 C; 479 G; 293 T; 1 other:
XX
alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638
alignment_block:
US-09-899-513-2 x AAA35309 ..
Align seg 1/1 to: AAA35309 from: 1 to: 1534
40 ProAlaArgPheIleValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
228 CCCACCAGG.....CTGGTCCCGCCCTCTATGGGCTGCTGCTGCTG 271

```

```

56 ugiyleProAlaAsnValAlaAlaLeuAlaMetHeLeArgSerGIg 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 GGGGCTCCGGCCCAATGGCGTGGCGCTGGTGGCTGGCCACGACGAC 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 LyrArgLeuGIgAlaLeuAlaLeuLeuTyrLeuPheAsnLeuAlaLeu 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 CTGGGCTGGCCCTCCACCATGCTG.....CTGATGAACTCTGGAGTGGT 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
90 AspGIuPhePheThrLeuThrLeuGIuLeuThrLeuThrLeuTyrLeu 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 GACCTCCTGTGGCGCCCTGGCGCTGGCCCCCGCGATGCGCTTACCACCTGGC 415
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 .....GlyLeuAlaArgArgProProAlaThrArgProGlyPro 118
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 TGGCCACGCGTGGCGCTTGGGGAGGCGCGCTGGCCGCTGGCCACGCGCG 465
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 ProThr.....ThrcysProPheMetArg 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
466 CACTGTGTGTCACATGATGATGCGTCATGTCGTGCTGTGGCCGCGCGACG 515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
127 gTrp.SerSerProArgSerSerAlaCysAlaAlaAlaSerTyrAla 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 CTGATGTGCATCCGCGGCGCTGTGACCCGCGCGGCGCGCGCGCTGTGC 565
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 TGGCGGGCGCGCTGGCCCTTGACATCTGCATGCTGCTGGCTCATGTGGCG 615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 .....ProGlyAlaTyrGI 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
616 CCGCGCTGGCACTGCCCTGCACATGCAGACGCGACACCTTCCGGCGTGGCG 665
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 ValaProArgAlaLeuProAla..... 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
666 CGCTCCGATCGCGTGTCTGCCATGACGCGCTGCCCTGGAGCGACAGCG 715
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 .....ProSerProGlyTrpArgAlaTrp..... 173
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
716 CTCCCACTGGCAACCGGCGCTTACCTGCTGGCGCTGTGGGCTGTTC 765
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 .....ProLeuProAlaTr 178
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
766 TGCCCGCTGTGGCGCATGCTGCTGTGCTACAGGCGCACACCTGCACACGCTG 815
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 pSerThrAlaGIgAlaAlaArgGIy..... 186
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
816 GCGCGCACCGCGCGCGCTACGCGCACGCGCTGAGCGTACCGCGACGTGT 865
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
187 .....TrpProProArgTrpProSerArg...ProProSerCysTrpCys 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
866 GCGTGGCTCCGCGCG...TGGCGCTTGTGTGGCCACGACACCTGTCTCTGC 912

```

seq_documentation_block:	
ID	AA32747 standard; cDNA; 4895 BP.
XX	
AC	AA32747;
XX	
DT	31-JAN-2000 (first entry)
XX	
DE	Human protease-activated receptor PAR4 cDNA.
XX	
KW	Protease-activated receptor; PAR4; ZCHEMR2; G protein coupled;
KW	cellular signalling; protease; tethered ligand; N-terminal;
KW	proteolysis; thrombin; trypsin; cleavage; hexapeptide; agonist;
KW	antagonist; cellular response; physiological response; clotting pathway
KW	platelet; proliferation; differentiation; mediation;
KW	inflammatory process; vascular injury; chemotaxis; mitogenesis;
KW	growth factor; production; ds.
XX	
OS	Homo sapiens.
XX	

Key	Location/Qualifiers
CDS	176..1333
5'UTR	/tag= a /product= "Human protease-activated receptor PAR4" 1..175
sig_peptide	/tag= b 176..226
mat_peptide	/tag= c 227..1330
3'UTR	/tag= d /product= "Mature human PAR4" 133..4895 /tag= e /note= "GC-rich"
WO950415-h2.	
07-OCT-1999.	
31-MAR-1999;	99WO-US07100.
01-APR-1998;	98US-0053866.
(Zymo) ZYMOGENETICS INC. (UNIV) UNIV WASHINGTON.	
Xu W, Presnell SR, Yee DP, Foster DC;	
WPI, 1999-633640/54.	
P-PSDB; AAY50135.	

PS Claim 1, Page 76-80; 85pp: English.

xx

cc This sequence represents a human protease-activated receptor PAR4 (also referred to as ZC6HMR) cDNA which was identified from EST (expressed sequence tag) sequences with homology to the three known protease-activated receptors (PAR1, PAR2 and PAR3). Protease-activated receptors (PARs) are a subfamily of G protein coupled receptors which are capable of mediating cellular signalling in response to proteases (e.g., thrombin). They are characterised by a tethered peptide ligand at the extracellular N-terminus that is generated by proteolysis. PAR is activated by thrombin or trypsin cleavage at Arg41/Gly48, which generates a new N-terminus corresponding to the tethered ligand (a hexapeptide). Agonists of PAR4 are useful for upregulating cellular or physiological responses whereas antagonists are used to downregulate these activities. The PAR4 protein is further useful for dissecting the effects of thrombin or other activating proteases in the clotting pathway from the effects of these proteases at the cellular level.

cc Agonists are specifically useful in promoting the proliferation and/or differentiation of platelets, in mediating inflammatory events, responses to vascular injury, chemotaxis or mitogenesis, and in producing growth factors. Antagonists are useful as research reagents for characterising sites of ligand-receptor interaction.

xx

SQ Sequence 4895 BP; 971 A; 1573 C; 1400 G; 951 T; 0 other;

alignment_scores:		
Quality:	177.50	length: 235
Ratio:	1.659	Gaps: 13
Percent Similarity:	45.532	Percent Identity: 30.638

alignment_block:
US-09-899-513-2 x AAZ32747

Align seg 1/1 to: AAZ32747 from: 1 to: 4895

```

40 ProLArgPheIleValProAlaIaTYraIaLeuAlaLeuGlyLe 56
|||::||| ::||| |||::|||::||| ::

```

```
401 CCCACGAG.....CTGTGCCCCCTCTATGGCTGTCTGTGGT 444
56 uGlyleuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGlyg 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 GGGGTGCGCCCAATGGCTGGCGCTGGGTGGTGGCCACGCGAGGAC 494
73 lYArgLeuGlyGlnAlaLeuLeuTyrLeuPheAsnLeuAlaVal 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CTGGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGCGACTGCT 538
90 AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuTyrLeu 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
539 GACCTCTGCTGGCGCTGGCGCTGGCGCGGATGGCTACACCTCGC 588
106 .....GlyleuAlaArgArgProProAlaThrArgProGlyPro 118
589 TGGCCAGCGCTGGCCCTTGGGAGGCGCGCTGGCGCGGCGCG 638
119 ProThr.....ThrcysProPromeTarArg 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
639 CACTCTATGTCACATGTATGGCTCAGTGTCTGTGGCGCGCTCAGC 688
127 gTTrp.SerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
689 CTGATGCGCTACCTGGCGCTGGCGACCGCGTGGCGCGCGCGCTGCG 738
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
739 TGGCGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTG 788
154 .....ProGlyAlaTyrGln 158
789 CCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 838
158 yAlaProArgAlaLeuProAla..... 165
839 CGCTCGATCGCTGGCTGGCGCTGGCGCTGGCGCTGGCGCTGG 888
166 .....ProSerProGlyTrpArgAlaTrp..... 173
889 CTCCACATGGCAACCGGCTTACCTGCTGGCGCTGGCGCTGGCT 938
174 .....ProLeuProAlaTrp 178
939 TGGCCCTGTGGCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 988
178 pSerThrAlaGlyGlnAlaArgGly..... 186
989 GCGGCGACGCGCGCGCTACGCGCGCTAGAGCGTACCGCAGTGGT 1038
187 ..TrrProProArgTrpProSerArg...ProProSerCysTrpCys 201
1039 GCTGGCTGTGGCG...TGGCTTCTTGTGGCGCAACACTGTGTGTGC 1085

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2000.DAT.AA250775
seq_documentation_block:
ID AA250775 standard; DNA; 4895 BP.
XX
XX AA250775;
AC
XX
XX 31-MAY-2000 (first entry)
DT
XX
XX Human protease activated receptor-4 DNA.
DE
XX
XX Human; PAR-4; protease activated receptor;
KW antiseptic molecule; PAR antibody; cytosolic; therapeutic;
KW metastatic tumour cell; placental implantation; invasive cell; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 176..1333
FT
```

```
FT /*tag= a
FT /product= "PAR-4"
XX
XX W0200008150-A1.
XX
XX 17-FEB-2000.
XX
XX 05-FEB-1999; 99WO-IL00079.
XX
XX 07-AUG-1998; 98IL-0125698.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Bar-Shavit R;
XX
XX WPI: 2000-205706/18.
XX
XX P-PSDB; AAY45036.
XX
XX Treating metastatic tumor cells useful for treating disorders involving
XX placenta implantation in a female comprises administration of an
XX antiseptic molecule complementary to an RNA sequence of a protease
XX activated receptor protein -
XX
XX Example 3; Fig 11a; 46pp; English.
XX
XX CC The patent discloses a method to treat metastatic tumour cells using
XX CC an antiseptic molecule comprising a polynucleotide complementary to an
XX CC RNA sequence of a protease activated receptor (PAR) protein, or an
XX CC antibody capable of binding to a PAR protein. The antiseptic molecules
XX CC and antibodies of PAR protein are also used to treat disorders associated
XX CC with implantation of placenta. The present sequence is a human PAR-4
XX CC DNA used for producing antiseptic molecules for treating invasive
XX CC cells.
XX
XX 50 Sequence 4895 BP; 971 A; 1573 C; 1400 G; 951 T; 0 other:

alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Caps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:
US-09-899-513-2 x AA250775 ..
Align seg 1/1 to: AA250775 from: 1 to: 4895
40 ProAlaArgPheIleValValProAlaAlaTyrAlaLeuAlaValGly 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 CCCACGAG.....CTGTGCCCCCTCTATGGCTGTCTGTGGT 444
56 uGlyleuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGlyg 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 GGGGTGCGCCCAATGGCTGGCGCTGGGTGGTGGTGGCCACGCGAGGAC 494
73 lYArgLeuGlyGlnAlaLeuLeuTyrLeuPheAsnLeuAlaVal 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CTGGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGCGACTGCT 538
90 AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuTyrLeu 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
539 GACCTCTGCTGGCGCTGGCGCTGGCGCGGATGGCTACACCTCGC 588
106 .....GlyleuAlaArgArgProProAlaThrArgProGlyPro 118
589 TGGCCAGCGCTGGCCCTTGGGAGGCGCGCTGGCGCGGCGCG 638
119 ProThr.....ThrcysProPromeTarArg 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
639 CACTCTATGTCACATGTATGGCTCAGTGTCTGTGGCGCGCTCAGC 688
127 gTTrp.SerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
689 CTGAGATGCTACCTGGCCCTGTGTGACACCGCTGGGCGCCGCTTCGCG 738
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
739 TGGCCGGGCGCTGGCCCTTGAGACTCTGATGGCTGTGGCTCATGGCCG 788
154 .....ProGlyAlaTyrG1 158
789 CCGGCGCTGGACCTGGCCCTGACACTGACGGGCGACAGACTTCGCGTGGCG 838
158 yAlaProArgAlaLeuProAla..... 165
839 CGGTCCGATCGCGCTGTGTGACATGACCGCGCTGCGCCGACGACAGCGC 888
166 .....ProSerProGlyTrpArgAlaTrp..... 173
889 CTCCCACTGGACACCGGCTTCACTGCTGGCGCTGTGGCGCTGTTCC 938
174 .....ProLeuProAlaTr 178
939 TGGCCCTGTGGCCATGCTGTGTGCTACGAGGGCGCACCTGACACAGCTG 988
178 pSerThrAlaGlyAlaAlaArgly..... 186
989 GCGGCCAGCGCGCGCTGACGCGCACGCGCTGAGGCTGACCGCAGTGTGT 1038
187 ..TTPProProArgTrpProSerArg...ProProSerCysTrpCys 201
1039 GCTGGCCCTCCGCG...TGGCCTTCTTGTGTGCGCCAGACACTGCTGCTGC 1085
```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH43632

seq_documentation_block:

ID AAH43632 standard; cDNA; 4925 BP.

AC AAH43632:

DT 21-JAN-2002 (first entry)

DE Human PAR4 coding sequence.

KM Human; protease-activated receptor; PAR4; N-terminal; activated PAR4;

KW platelet aggregation; inhibition; tumour cell; proliferation; ss.

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 191..1348

FT /tag= a

FT /product= "PAR4"

FT /note= "CDS derived from protein AAB47623"

PN MO200158930-A1.

PD 16-AUG-2001.

PF 06-FEB-2001; 2001WO-US03807.

PR 09-FEB-2000; 2000US-0500646.

PA (ZYMO) ZYMOGENETICS INC.

PI Balindur N, West RR;

DR WPI: 2001-656678/75.

DR P-PSDB; AAB47623.

PS Disclosure; Page 51-58; 84pp; English.

XX This sequence encodes human protease-activated receptor, PAR4. Peptides
CC derived from, or based on, the N-terminal of activated PAR4 (see
CC AAB47624-77), are capable of stimulating PAR4. These peptides may be
CC used to activate PAR4 at lower concentrations than wild type PAR4.
CC These peptides may be used to stimulate platelet aggregation, and for
CC inhibiting tumor cell proliferation.

SQ Sequence 4925 BP; 977 A; 1582 C; 1411 G; 955 T; 0 other;

alignment_scores:

Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:

US-09-899-513-2 x AAH43632.

Align seg 1/1 to: AAH43632 from: 1 to: 4925

```
40 ProAlaTrpPheLeuValProAlaTyrAlaLeuAlaLeuGlyLe 56
416 CCGACACAGG.....CTGGTCCCGCCCTTATGGGCTGTCTGTGTGT 459
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGly 73
460 GGGGCTGGCGGCAATGGGCTGGCGCTGTGGCTGTGGCAGCAGGAC 509
73 lYArgLeuGlyAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
510 CTGGGCTCCCTCCACCATGCTG.....CTGATGAACCTGCGACACTCT 553
90 AspGluPhePheThrLeuThrLeuGlnLeuTrpLeuThrTyrTyrLeu 105
554 GACCTCTGTGGCGCTGGCCCTGCGCCGCGGATGCGCTACCACTGCG 603
106 .....GlyLeuAlaArgProProAlaTrpArgProGlyPro 118
604 TGGCAGCGCTGGCCCTTGGGAGGCGCGCTGCGCTGGCCAGCGCG 653
119 ProThr.....ThCysProMetArgArg 127
654 CACTTATGTCATCATGTATGCTCAGTGTGCTGGCGCGCTCAGC 703
127 gTrp.SerSerProArgSerSerAlaCysAlaAlaAlaSerTyrAla 143
704 CTGATCGCTACCTGGCCCTGTGTGACCGCTGGCGCGCGCGCTGCG 753
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
754 TGGCCGGGCGCTGGCCCTTGAGACTCTGATGATGCTGTGGCTCATGGCG 803
154 .....ProGlyAlaTyrG1 158
804 CCGGCGCTGGACCTGGCCCTGACACTGACGCGACAGACTTCGCGTGGCG 853
158 yAlaProArgAlaLeuProAla..... 165
854 CGTCCGATCGCGCTGTGTGACATGACCGCGCTGAGCGCACAGCGC 903
166 .....ProSerProGlyTrpArgAlaTrp..... 173
904 CTCCCACTGGACACCGGCTTCACTGCTGGCGCTGTGGCGCTGTTCC 953
174 .....ProLeuProAlaTr 178
954 TGGCCCTGTGGCCATGCTGTGTGCTACGAGGGCGCACCTGACACAGCTG 1003
178 pSerThrAlaGlyAlaAlaArgly..... 186
1004 CCGGCGACGGCGCGCTACGGCCAGCGCTGAGGCTGACCGCAGTGTGT 1053
```

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'universal' or alternative base (I') can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiallergic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and/or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system transmitters, CNS and peripheral nervous and non-nervous system transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (cf), allergic rhinitis (AR), pulmonary

107 ..11P10P10a1g11P10se1a1g...1P10se1c1y1s1r1p1c1y1s 201

XX 22-NOV-1999 (first entry)
 XX
 DE Human G protein coupled receptor GPR25 cDNA.
 XX
 KW GPR25; G protein coupled receptor; human; cancer; infection;
 XX pain; diagnosis; therapy; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 79..1164
 FT /*tag= a
 XX
 PN MO9945030-A1.
 XX
 PD 10-SEP-1999.
 XX
 XX 01-MAR-1999; 99MO-US04424.
 XX
 PR 16-MAR-1998; 98US-0039608.
 XX 04-MAR-1998; 98US-0034985.
 XX
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 XX
 PI Elshourbagy N, Lane PA, Tsui P;
 DR WPI; 1999-540815/45.
 DR P-PSDB; AAY31751.
 XX
 PT New messenger polypeptides and polynucleotides useful for diagnosis,
 PT prevention and treatment of cancer and HIV infection - and host
 PT cells and antibodies
 XX
 PS Claim 2; Page 29; 42pp; English.
 XX
 CC This is the nucleotide sequence of an isolated polynucleotide
 CC encoding a novel human G protein coupled receptor termed GPR25
 CC (see AAY31751). The polynucleotide can be obtained from a cDNA
 CC library derived from mRNA in cells of human spleen using expressed
 CC sequence tag analysis. Treatment of a subject in need of GPR25
 CC comprises administering GPR25, or a polynucleotide encoding a
 CC polypeptide that effects production of GPR25 in vivo, while
 CC treatment of a subject in need of inhibiting activity or expression
 CC of GPR25 comprises administering an antagonist of GPR25, a nucleic
 CC acid that inhibits expression of GPR25, or a polypeptide that
 CC competes with GPR25 for its ligand, substrate or receptor. Also
 CC claimed are an expression system, host cells, a membrane of the
 CC host cells and preparation of GPR25. GPR25 is useful for diagnosing
 CC and treatment of a disease or susceptibility to a disease by
 CC determining the presence or absence of a mutation in GPR25 nucleic
 CC acid, and/or analyzing for the presence or amount of GPR25
 CC expression (claimed). It is also useful for identifying compounds
 CC that stimulate or inhibit GPR25, by measuring binding of the
 CC candidate to GPR25, and analyzing signal of GPR25 (claimed).
 CC Expression in vivo is useful in gene therapy for treatment of
 CC disease. Diseases that can be treated or diagnosed include fungal,
 CC protozoal and viral (e.g. HIV) infections, pain, cancer, diabetes,
 CC obesity, anorexia, bulimia, asthma, Parkinson's disease, acute
 CC heart failure, hypotension, hypertension, urinary retention,
 CC osteoporosis, angina pectoris, myocardial infarction, stroke,
 CC ulcer, allergy, benign prostatic hypertrophy, migraine, vomiting,
 CC anxiety, schizophrenia, manic depression, delirium, dementia,
 CC Huntington's disease and Gilles de la Tourette syndrome.
 XX
 SQ Sequence 1174 BP; 116 A; 443 C; 408 G; 207 T; 0 other;

alignment_scores:
 Quality: 167.50 Length: 200
 Ratio: 1.565 Gaps: 10
 Percent Similarity: 53.500 Percent Identity: 34.500

alignment_block:
 US-09-899-513-2 x AAX87944 ..
 Align seg 1/1 to: AAX87944 from: 1 to: 1174
 24 SerSerValAlGlySplenglyValGly..Gly.....C 34
 117 ACGCGCTGAGACTACTCGGGGTTGAGCGCTGAGAGAGCTGAGCTGT 166
 34 ySerLeuTrpAspAspProAlaArgPheIleValValProAlaAlaTyr 50
 167 GTCCGGCGCGGAGCTGCGCTACGCTGCTACCTACCTCCCGGCTCTAC 216
 51 AlAlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAla 67
 217 CTGGCGGCGCTTGGCGCTGCGCTGCTGCGCAAC.....GCTTTGTGTG 260
 67 tPheIleArgSerGlyGlyArgLeuGlyGlnAlaLeuLeu...LeuTyrL 83
 261 GTGGCTGCTGGCGCGGCGGCGGCGGCGGCGGCTGTGATACCTTGG 310
 83 euphAsnLeuAlaLeuValAspGluPhePheThrLeuThrLeuGlnLeu 99
 311 TCGTCACCTGGCGGCGAGCTGAGCTGCTGTGCTGCTGCTGCTGCTGCTG 360
 100 TrpLeuThrTyrTyrLeuGlyLeuAlaArgArgProAlaThrArgPr 116
 361 TGGGCGCGCGG.....GCGGCGCTAGGCGGCGGCGGCGGCTGCGG 403
 116 oGlyProProThrThrCysProPheMetArgTyrSerSerProArgS 133
 404 ATGG...CCTGTCAGAGCTGAGAGCTTGGCGCTGCGGCGGCGGCTGCG 450
 133 eSerAlaCysAlaAlaAlaAlaSerTyrAlaValProGlyProGlyArg 149
 451 GCGGCGCGCTGCTGCTGCTGCGGCGGCGGCGGCGGCGGCTGCGCTGCG 500
 150 LeuProAlaTrpProGly..AlaTyrGlyAlaProAlaAlaLeu..... 163
 501 GTGAAAGCTGCTGAGGCGGAGGCGGAGGCGGAGGCGGCGGCGGCGGCTG 550
 164ProAlaProSerProGlyTyrArgAlaTrpPro...LeuProAl 177
 551 CCGGCTGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCTGCG 600
 177 aTrpSerThrAlaGly.....GlnAlaArg 186
 601 CTGCTTACCGGGGCTTGCAGCCCTGCTGCGGCGGCGGAGAGCAGCAGTG 650
 186 lYTrpProProArgTyrTrpProSerArgProProSerCysTrpCys 201
 651 CGGCGAGAGCCCTCCACGCGCTTCCAGGCGCTCAGCTGCTGCTGCTG 697
 seq_name: /SID51/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA10298
 seq_documentation_block:
 ID AAA10298 standard; DNA; 1464 BP.
 XX
 AC AAA10298;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE DNA encoding human APJ receptor.
 XX
 KW APJ receptor; HIV-1 coreceptor; SIV; cellular infection;
 XX envelope protein; env; CD4 coexpression; drug development; ds.
 OS Homo sapiens.
 XX
 FH key Location/Qualifiers
 FT CDS 199..1341
 FT /*tag= a
 FT /*product= "Human APJ receptor"


```

XX MN09801468-A1.
XX PD 15-JAN-1998.
XX PF 30-JUN-1997; 97WO-US11311.
XX PR 05-JUL-1996; 96US-0021672.
XX PA (UYRP ) UNIV ROCHESTER.
XX PI Reynolds PR:
XX P1
XX DR WPI: 1998-110224/10.
XX DR P-PSDB; AAM52285.
XX PT
XX PT New isolated chondrocyte protein - which is involved in the
XX PT transition of chondrocytes from proliferation to hypertrophy, used
XX PT to develop products for diagnosis and therapy
XX PT
XX Claim 15; Page 61; 85pp; English.
XX
XX The sequence is that encoding a protein which is selectively
XX expressed in chondrocytes in lower proliferative or upper
XX hypertrophic zones of long bone and embryonic vertebrate growth
XX plates. The protein is involved in the transition of
XX chondrocytes from proliferation to hypertrophy. The product
XX can be used for identifying the occurrence of proliferation
XX or hypertrophy or the transition of chondrocytes from proliferation
XX to hypertrophy in a tissue sample. It can also be used for
XX preventing chondrocytes from transitioning from proliferation
XX to hypertrophy and for inhibiting arthritic progression of
XX articular chondrocytes in a patient. It can also be used for
XX inducing chondrocytes to transition from proliferation to
XX hypertrophy and for treating non-union bone defects.
XX
XX Sequence 954 BP; 157 A; 330 C; 308 G; 159 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 162.50      Length: 129
XX      Ratio: 2.621      Gaps: 7
XX Percent Similarity: 48.062 Percent Identity: 34.109
XX
XX alignment block:
XX US-09-899-513-2 x AAV19862 ..
XX
XX align seg 1/1 to: AAV19862 from: 1 to: 954
XX
XX 110 ATGProProAlaThrArgProGlyProProThrhrcysProPometar 126
XX      111      ::::: 111      ||||| 111111111
XX 125 CGCTGTCCTCGGGCCGACGCGCCGCCGCCGCTGCCGCCG... 169
XX
XX 126 gArGTrPserSerProArg...SerSerAlaCysAlaAlaAlaSerT 142
XX      170 .....CTGCGCCCGCTGACGCTCGCGCGCTCGTCGCGCAC 212
XX
XX 142 yRAlaValProGlyProGlyArgLeuProAlaTPro.....Gly 155
XX      213 CGCGTACCCACGCGCGGGCAATCCGCGCTGGCCGACGTGCACGGCC 262
XX
XX 156 AlaTyGlyAlaProArgAlaLeuProAlaProSerPro..... 168
XX      263 GCGTCGCGCGCGCGCGCGCGCTGCCCGCGCGCGCGCGTGGCC 312
XX
XX 169 .....GlyTrPArgAlaT 173
XX
XX 313 GCCTCGCCGATGTGTAAGAGAGAGCTGCACGGCGCGCGCGCG 362
XX
XX 173 rProLeuProAlaTrpSerThrAlaGlyAlaArgLYTrpProPro 189
XX      363 GGGCGCGCGCGACATGAAACACCTGGCGCGCGCC.....TGGCGCGCC 406
XX
XX 190 Pro.....ArgTrpProSerAr 195

```

```

407 GCTTCCCGCGCTTTCCGCCGCCGCCGCTCGGCGCTGGCCAGCAGC 456
|||||
195 gPro.....ProSerCyStPcysSerArgProThr 205
|||||
457 TCCAAGCACCCGCTGCCTCAGACGCGCGGCGCTTCC 493

seq_name: /SIBS1/gcdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19864

seq_documentation_block:
ID   AAV19864 standard; DNA; 1347 BP.
XX
XX   AAV19864;
XX
XX   21-JUL-1998 (first entry)
XX
XX   Gallus domesticus chondrocyte protein gene.
DE
XX   chondrocyte; transition;; proliferation; hypertrophy; diagnosis;
XX   therapy; long bone; lower proliferative; upper hypertrophic;
XX   embryonic vertebrae growth plates; arthritis; treatment;
XX   non-union bone defects; ss.
XX
XX   Gallus domesticus.
OS
FH   Key Location/Qualifiers
FT   CDS 1..1347
FT     /*tag= a
FT     /product= chondrocyte protein
FT
XX   MO9801468-A1.
PN
XX   15-JAN-1998.
PD
XX   30-JUN-1997; 97WO-US11311.
PF
XX   05-JUL-1996; 96US-0021672.
PR
XX   (UTRP ) UNIV ROCHESTER.
PA
XX   Reynolds PR;
PI
XX   WPI: 1998-110224/10.
DR   P-PSDB: AAW52286.
XX
XX   New isolated chondrocyte protein - which is involved in the
XX   transition of chondrocytes from proliferation to hypertrophy, used
XX   to develop products for diagnosis and therapy
XX
XX   Claim 19; Pages 64-65; 85pp; English.
XX
XX   The sequence is that encoding a protein which is selectively
XX   expressed in chondrocytes in lower proliferative or upper
XX   hypertrophic zones of long bone and embryonic vertebrae growth
XX   plates. The protein is involved in the transition of
XX   chondrocytes from proliferation to hypertrophy. The product
XX   can be used for identifying the occurrence of proliferation
XX   or hypertrophy in a tissue sample. It can also be used for
XX   to hypertrophy in a tissue sample. It can also be used for
XX   preventing chondrocytes from transitioning from proliferation
XX   to hypertrophy and for inhibiting arthritic progression of
XX   articular chondrocytes in a patient. It can also be used for
XX   inducing chondrocytes to transition from proliferation to
XX   hypertrophy and for treating non-union bone defects.
XX
SQ   Sequence 1347 BP; 266 A; 432 C; 392 G; 257 T; 0 other;

alignment_scores:
Quality: 162.50 Length: 129
Ratio: 2.621 Gaps: 7
Percent Similarity: 48.062 Percent Identity: 34.109

```



```

XX DT 21-JUL-1998 (first entry)
XX XX Gallus domesticus chondrocyte protein gene.
XX DE
XX KM chondrocyte; transition; proliferation; hypertrophy; diagnosis;
XX KM therapy; long bone; lower embryonic vertebrae growth
XX KM embryonic vertebrae growth plates; arthritis; treatment;
XX KM non-union bone defects; ss.
XX OS
XX Gallus domesticus.
XX Key Location/Qualifiers
XX FT CDS 81.1037
XX FT /tag a
XX FT /product= chondrocyte protein
XX PN
XX MO9801468-A1.
XX 15-JAN-1998.
XX PF
XX 30-JUN-1997; 97WO-DS1311.
XX XX
XX PR 05-JUL-1996; 96US-0021672.
XX XX
XX (UYRP ) UNIV ROCHESTER.
XX PA
XX Reynolds PR;
XX XX
XX WP1; 1998-110224/10.
XX DR P-PSDB; AAW52285.
XX XX
XX New isolated chondrocyte protein - which is involved in the
XX PT transition of chondrocytes from proliferation to hypertrophy, used
XX PT to develop products for diagnosis and therapy
XX PT
XX Claim 14; Pages 59-60; 85pp; English.
XX PS
XX XX The sequence is that encoding a protein which is selectively
XX CC expressed in chondrocytes in lower proliferative or upper
XX CC hypertrophic zones of long bone and embryonic vertebrae growth
XX CC plates. The protein is involved in the transition of
XX CC chondrocytes from proliferation to hypertrophy. The product
XX CC can be used for identifying the occurrence of proliferation
XX CC or hypertrophy or the transition of chondrocytes from proliferation
XX CC to hypertrophy in a tissue sample. It can also be used for
XX CC preventing chondrocytes from transitioning from proliferation
XX CC to hypertrophy and for inhibiting arthritic progression of
XX CC articular chondrocytes in a patient. It can also be used for
XX CC inducing chondrocytes to transition from proliferation to
XX CC hypertrophy and for treating non-union bone defects.
XX CC
XX Sequence 2233 BP; 534 A; 579 C; 601 G; 519 T; 0 other;
XX SO
XX
XX Alignment_scores:
XX Quality: 162.50 Length: 129
XX Ratio: 2.621 Gaps: 7
XX Percent Similarity: 48.062 Percent Identity: 34.109
XX
XX Alignment_block:
XX US-09-899-513-2 x AAV19861 ..
XX
XX Align seg 1/1 to: AAV19861 from: 1 to: 2233
XX
XX 110 ArgProPoaIaThrArgProGlyProProThrThrcysProBroweTar 126
XX ||| |||::: ||| ||||| ||||| ||||| |||||
XX 205 CGCTGTCCTCGCGGCGGACGCCGCCGCCGCCGCGCTGCGCGCG... 249
XX
XX 126 gArGrTrpSerProArg...SerSerAlaCysAlaAlaAlaAlaSerT 142
XX ::||| ||||| ||||| ||||| ||||| ||||| |||||
XX 250 .....CCTGCGCGCGGCTGACCTCGCGCGGCTGCTCGGCACGCGCAC 292

```

```

142 yfAlaValaProGlyProGlyArgLeuProAlaTrpPro..... Gly 155
    |||::||| |||||::||| |||||::|||
293 CGCTACCCACGCGCGGCAATCCGCGCTGCGAGCTGCACGCGC 342
    ||::||| ||||| |||
343 GCCTCGCGCGCGCGCGCGCGCTCTGCGCGCGCGCGCGCTGCGC 392
    |||
169 ..... GlyTrpArgAlaTr 173
    |||||
393 GCCTGCGCGATGTGTACGAGAGACCTCGACGCGCGCTGCGCGCGC 442
173 rpProLeuProAlaTrpSerThrAlaGlyGlnAlaArgGlyTrpProPro 189
    ::|||::||| |||
443 GGGCGCGCGCGACATGGAACACCTGGCGCGCGC..... TGCGCGCGC 486
190 Pro..... ArgTrpProSerAr 195
    |||||
487 GCTTCCCGCGCTCTTCCGCGCGCGCGCGCGCTTGGCGCTGGCCACGACG 536
195 gPro..... ProSerCysTrpCysSerArgProThr 205
    ||::||| ::|||::|||
537 TCCAAGCACCGCTGCTGCGAGCGCGCGCGCTTCC 573

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/MA1998.DAT:AAV19860
seq_documentation_block:
ID AAV19860 standard; DNA: 5027 BP.
XX
XX AAV19860;
XX
XX 21-JUL-1998 (first entry)
XX
XX Gallus domesticus chondrocyte protein gene.
XX
XX chondrocyte; transition;; proliferation; hypertrophy; diagnosis;
XX therapy; long bone; lower proliferative; upper hypertrophic;
XX embryonic vertebrae growth plates; arthritis; treatment;
XX non-union bone defects; ss.
XX
XX Gallus domesticus.
XX
XX OS
XX
XX FT FT Location/Qualifiers
XX CDS 81..1037
XX /tag=a
XX /product= chondrocyte protein
XX
XX WO9801468-A1.
XX
XX 15-JAN-1998.
XX
XX 30-JUN-1997; 97WO-0611311.
XX
XX 05-JUL-1996; 96US-0021672.
XX
XX (UYRP ) UNIV ROCHESTER.
XX
XX Reynolds PR;
XX
XX WIPI: 1998-110224/10.
XX P-PSDB: AAW52285.
XX
XX New isolated chondrocyte protein - which is involved in the
XX transition of chondrocytes from proliferation to hypertrophy, used
XX to develop products for diagnosis and therapy
XX
XX Claim 13; Pages 55-58; 85pp; English.
XX
XX The sequence is that encoding a protein which is selectively
XX expressed in chondrocytes in lower proliferative or upper
XX hypertrophic zones of long bone and embryonic vertebrae growth
XX plates. The protein is involved in the transition of
XX chondrocytes from proliferation to hypertrophy. The product

```

CC can be used for identifying the occurrence of proliferation
CC or hypertrophy or the transition of chondrocytes from proliferation
CC to hypertrophy in a tissue sample. It can also be used for
CC preventing chondrocytes from transitioning from proliferation
CC to hypertrophy and for inhibiting arthritic progression of
CC articular chondrocytes in a patient. It can also be used for
CC inducing chondrocytes to transition from proliferation to
CC hypertrophy and for treating non-union bone defects.
XX

Sequence 5027 BP: 1353 A: 1136 C: 1200 G: 1338 T: 0 other:

alignment_scores: Quality: 162.50 Length: 129
 Ratio: 2.621 Gaps: 7
Percent Similarity: 48.062 Percent Identity: 34.109

alignment_block:
US-09-899-513-2 x AAV19860 ..

align seg 1/1 to: AAV19860 from: 1 to: 5027

```
110 ArgProProAlaThrArgProGlyProProThrThrcysProPrometar 126
    ||| |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 CGCTGCTCCCTCGGCGCCACGCCGCCGCCGCCGCTGCGCCGCG... 249
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 gArgTrpSerSerProArg...SerSerAlaCysAlaAlaAlaIasert 142
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 .....CTGCGGCCCGCTGCACCTCGCGCGCGCTGCGCGCAC 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 yAlaValProGlyProGlyArgLeuProAlaTrpPro.....Gly 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 CCCTACCCACGCGCGCGCAATCCGCGCGCTGCGCGAGCTGCACGCGCC 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 AlaTrpGlyAlaProArgAlaLeuProAlaProSerPro..... 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 GCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCC 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 .....GlyTrpArgAlaTr 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 GCCTGCGCGATGTGTACGAGAGAGCTCGACGCGCGCTGCGCGCGCG 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 rProLeuProAlaTrpSerThrAlaGlyAlaAlaArgGlyTrpProPro 189
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 GGGCGCGCGACATGAGAACACTGCGCGCGCGCGCGCGCGCGCGCG 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 Pro.....ArgTrpProSerAr 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
487 GCTTCCCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCTGCGCGCGAC 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 gPro.....ProSerCysTrpCysSerArgProThr 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
537 TCCAAGCACCGCTGCTGACAGAGCGGCGGCGCTTCC 573
```



```

|||||
382 ACSTGGCGCCCTGCAATGTCATCCAGCGCGCGCTGGCCAG 333
78 AAlaLeuLeuTyrLeuPheAsnLeuValAspGluPhePhe 94
332 GCCCTGCTTCTACTCTCAACTGCTGTTGATGATCTTTCAC 283
94 rLeuThrLeuGlnLeuTyrPleuThrTyrTyrLeuGlnLeuAlaArgArg 111
282 GCTACGCTGACGCTGTGCTCCTACTACTGCGCTGCGCCGCGAGGC 233
111 rGProAlaThrArgProGlyProProThrThrCysProPheArgArg 127
232 CGCCTGCCACGCGCGCGGCGCACCTACTACTGTCACCTATGCGCGCG 183
128 TrpSerSerProArgSerSerAlaCysAlaAlaAlaSerTyrAlaVal 144
182 TGGTCTTGGCGCGCTCATACAGCTGTGCGCTGCGCTTCTTACGCGGT 133
144 lProGlyProGlyArgLeuProAlaTyrProGlyAlaTyrGlyAlaPro 161
132 CCGCGGCGCCAGGCGCGCTGCCCGCTGCGCTGCTACGCGCGCGCC 83
161 rGAlaLeuProAlaProSerProGlyTyrPargAlaTyrProLeuProAla 177
82 GCGCGCTGCTGCTGCTGCTGCTGCGCGCGCTGCGCGCTGCTGCTGCC 33
178 TrpSerThrAlaGlyAlaAlaArgGlyTyr 187
32 TGGACACCGCTGGCGCAAGCTGGGGCTGG 3

```

seq_name: gb_est2:BF195670

seq_documentation_block:

LOCUS BF195670 490 bp mRNA linear EST 03-NOV-2000

DEFINITION 7186906.x1 NCI-CGAP-Ov18 Homo sapiens CDNA clone IMAGE:3571762 3'

ACCESSION BF195670

VERSION BF195670.1 GI:11082804

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

MENT Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaaps-remail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emerit-Buck, M.D., Ph.D. CDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL, send email to:

info@image.lnl.gov

High quality sequence stop: 452.

Location/Qualifiers

1. 490

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3571762"

/clone_id="NCI-CGAP-Ov18"

/tissue_type="fibrocyte"

seq_name: gb_est1:A1264420

seq_documentation_block:

LOCUS A1264420 460 bp mRNA linear EST 28-JAN-1999

DEFINITION 0K54F04.x1 NCI CGAP C08 Homo sapiens CDNA clone IMAGE:1872799 3'

ACCESSION A1264420

VERSION A1264420.1 GI:3872623

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

MENT Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaaps-remail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emerit-Buck, M.D., Ph.D. CDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL, send email to:

info@image.lnl.gov

High quality sequence stop: 452.

Location/Qualifiers

1. 490

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3571762"

/clone_id="NCI-CGAP-Ov18"

/tissue_type="fibrocyte"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pTZ19-Pac (Pharmacia) with a

modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGCGACATTTTTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pTZ19 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Patricia Bonaldo.

alignment_scores:

Quality: 856.00 Length: 158

Ratio: 5.452 Gaps: 0

Percent Similarity: 99.367 Percent Identity: 98.734

alignment_block:

US-09-899-513-2 x BF195670/rev ..

Align seg 1/1 to reverse of: BF195670 from: 1 to: 490

39 ASPProAlaArgPheIleValAlProAlaAlaTyrAlaLeuAlaLeu 55

488 GACCGCTGCTGCTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 439

55 YLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSer 72

438 CCGGGGGCTGCCAGCAACGTCAGCGCCCTGCGCAATGTCATCCAGCG 389

72 YGlyArgLeuGlyGlnAlaLeuLeuTyrLeuPheAsnLeuAlaLeu 88

388 GGGGGCGCTGGGCCAGCGCGCTGCTTCTTACTGTTCACTGGGCTTG 339

89 ValAspGluPhePheThrLeuThrLeuGlnLeuTyrPleuThrTyrTyr 105

338 GTTGAATGAATTTCTTCACGCTACGCTGCTGCTGCTGCTGCTGCT 289

105 uGlyLeuAlaArgArgProProAlaThrArgProGlyProProThrTh 122

288 GGGCGCTGGCGCGAGCGCGCTGCGCACGCGCGCGCGCGCGCTACT 239

122 VaProPheMetArgArgTrpSerSerProArgSerSerAlaCysAla 138

238 GTCCACTATGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 189

139 AlaAlaSerTyrAlaValProGlyProGlyArgLeuProAlaTyrPro 155

188 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 139

155 ValArgGlyAlaProAlaGlyAlaLeuProAlaProSerProGlyTyr 172

138 TGCCTACGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 89

172 lAtTrpProLeuProAlaTyrPheThrAlaGlyAlaAlaArgGlyTyr 188

88 CCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 39

189 ProProArgTrpProSerArgPro 196

38 CGCCACAGGTGGCTTGGCGCGCG 15

REFERENCE 1 (bases 1 to 460)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1000 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 449.
Location/Qualifiers
1. 460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1872799"
/clone_lib="NCI-CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo.
BASE COUNT 87 a 151 c 179 g 42 t 1 others
ORIGIN
alignment_scores:
Quality: 816.00 Length: 153
Ratio: 5.368 Gaps: 0
Percent Similarity: 99.346 Percent Identity: 98.693
alignment_block:
US-09-899-513-2 x AI264420/rev ..
Align seg 1/1 to reverse of: AI264420 from: 1 to: 460
38 ASPAPPPALAArGPhellValProAlaAlaTyrAlaLeuAlaLe 54
|||||
460 GATGACCTGCTGCTTCATCGTGGCCGCGGCTATGCTTGGCACT 411
54 uGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgS 71
|||||
410 GGGCGTGGGGCTCCAGCCAACTGNCGGCGCTGCAATGTCATCCGCA 361
71 erGlyGlyArgLeuGlyGlnAlaLeuLeuTyrLeuPheAsnIleAla 87
|||||
360 GCGGCGGGCGCTGGGGCAGGCCCTGCTTCACTGTTCAACCTGGCT 311
88 LeuValAspGluPhePheThrLeuThrLeuGlnLeuTyrPheTyrTy 104
|||||
310 CTGGTTGATGAGTTCCTTCAAGCTCAGCTGCTGCTGCTACACTA 261
104 rLeuGlyLeuAlaAlaArgProProAlaThrArgProGlyProProThrT 121
|||||
260 CCTGGGCTGGCCCGGAGGCCCTGCTCAGCGCGCGGCGCCACCTACTA 211
121 hrcysProPomeTArgArgTirPserSerProArgSerSerAlaCysAla 137
|||||
210 CGTGTCCACTATGCGGGGTGGTCTTGGCGGCTCATAGGGGTGGCC 161
138 AlaAlaAlaSerTyrAlaValProGlyArgLeuProAlaTirPpr 154

|||||
160 GCTGCGGCTTCGTACCGGCGGCCGAGCGGCGCTGCCCTGGGCC 111
154 oGlyAlaTyrGlyAlaProAlaArgAlaLeuProAlaProSerProGlyTyr 171
|||||
110 CGGTGCTTACGGCGGCCCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 61
171 rGAlaTrrProLeuProAlaTrrSerThrAlaGlyGlnAlaArgGlyTyr 187
60 GGGCTTGGCCCCCTTCTGCTTGAAGCACCCTGGCAAGCTCGGGCTGG 11
188 ProProPro 190
|||||
10 CTCCGCCCA 2
seq_name: gb_est1:AI277080
seq_documentation_block:
LOCUS AI277080 466 bp mRNA linear EST 29-JAN-1999
DEFINITION q140h02.x1 NCI-CGAP_C08 Homo sapiens CDNA clone IMAGE:1874835 3',
mRNA sequence.
ACCESSION AI277080
VERSION AI277080.1 GI:3899348
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 978 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 431.
Location/Qualifiers
1. 466
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1874835"
/clone_lib="NCI-CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo.
BASE COUNT 89 a 153 c 177 g 45 t 2 others
ORIGIN
alignment_scores:
Quality: 805.00 Length: 153
Ratio: 5.331 Gaps: 0
Percent Similarity: 98.693 Percent Identity: 98.039
alignment_block:
US-09-899-513-2 x AI277080/rev ..

Align seg 1/1 to reverse of: A1277080 from: 1 to: 466

```

34 CysSerLeuTrpAspProAlaArgPheIleValProAlaIaIaTy 50
464 TGCAGCCTCTGGNATGACCTGCTGCTTCATCGTGGTCCCGGCGCTTA 415
50 rAlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaIaLeuAlaIa 67
414 TGCCTGGCAGCTGGGCGCTGGGCGCTGCCAGCAGCTGCGGCGCTGCA 365
67 etPheIleArgSerGlyArgLeuGlyGlnAlaLeuLeuLeuTyTru 83
364 TGTTCATCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCA 315
84 PheAsnLeuAlaLeuValAspGluPhePheTruLeuTruLeuGlnLeuTr 100
314 TTCAACCTGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 265
100 PleuThrTyTruLeuGlyLeuAlaAlaArgArgProAlaIaIaArgProG 117
264 GCTACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 215
117 TyProThrTyTruLeuGlyLeuAlaAlaArgArgProAlaIaIaArgProG 133
214 GGCCACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 165
134 SerAlaCysAlaAlaAlaAlaAlaSerTyAlaValProGlyProGlyArgLe 150
164 TCAGCGTGTGCGGCGCTGCGGCTGCTGACGCGGCTGCGGCGGCGGCGG 115
150 uProAlaTrpProGlyAlaTyGlyAlaProArgAlaLeuProAlaIaProS 167
114 GCCCGCTGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65
167 etProGlyTrpArgAlaTrpProLeuProAlaIaTrpSerThrAlaGln 183
64 CGCCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 15
184 AlaArgGly 186
14 GCTCGGGGCG 6

seq_name: gb_est2:BF000091
seq_documentation_block:
LOCUS BF000091 449 bp mRNA linear EST 06-OCT-2000
DEFINITION 7h17d11.x1 NCI-CGAP_Col6 Homo sapiens CDNA clone IMAGE:3316245 3'
similar to contains PTR5.t3 TARI repetitive element ;, mRNA
sequence.
ACCESSION BF000091
VERSION BF000091 GI:10700366
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 449)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Info@image.lnl.gov
Seq primer: -400p from Glibco

```

High quality sequence stop: 388.
Location/Qualifiers
1. 449

```

FEATURES
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3316245"
        /clone_lib="NCI-CGAP_Col6"
        /tissue_type="colon tumor, RER+"
        /lab_host="DH10B"
        /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
        modified polylinker; Site_1: Not I; Site_2: Eco RI;
        Plasmid DNA from the normalised library NCI CGAP Col6 was
        purified, and ss circles were made in vitro. Following HAP
        hybridization reaction. The driver was PCR-amplified cDNAs
        from a pool of 5,000 clones made from the same library
        (clonids 1057416-1061255, and 114584-1145351).
        Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      87 a      150 c      172 g      40 t
ORIGIN

```

alignment_scores: Quality: 802.00 Length: 149
Ratio: 5.419 Gaps: 0
Percent Similarity: 99.329 Percent Identity: 99.329

alignment_block:
us-09-899-513-2 x BF000091/rev ..

Align seg 1/1 to reverse of: BF000091 from: 1 to: 449

```

34 CysSerLeuTrpAspProAlaArgPheIleValProAlaIaIaTy 50
447 TGCAGCCTCTGGNATGACCTGCTGCTTCATCGTGGTCCCGGCGCTTA 398
50 rAlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaIaLeuAlaIa 67
397 TGCCTTGGCAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 348
67 etPheIleArgSerGlyArgLeuGlyGlnAlaLeuLeuLeuTyTru 83
347 TGTTCATCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 298
84 PheAsnLeuAlaLeuValAspGluPhePheTruLeuTruLeuGlnLeuTr 100
297 TCACACCTGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 248
100 PleuThrTyTruLeuGlyLeuAlaAlaArgArgProProAlaIaIaArgProG 117
247 GCTCACCTACTACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 198
117 TyProThrTyTruLeuGlyLeuAlaAlaArgArgProProAlaIaIaArgProG 133
197 GCCACCTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 148
134 SerAlaCysAlaAlaAlaAlaAlaSerTyAlaValProGlyProGlyArgLe 150
147 TCAGCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 98
150 uProAlaTrpProGlyAlaTyGlyAlaProArgAlaLeuProAlaIaProS 167
97 GCCCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 48
167 etProGlyTrpArgAlaTrpProLeuProAlaIaTrpSerThrAlaGln 182
47 CGCCTGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1

seq_name: gb_est2:BF108800
seq_documentation_block:
LOCUS BF108800 392 bp mRNA linear EST 20-OCT-2000
DEFINITION 7152h02.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens CDNA clone

```

IMAGE:3525338 3', mRNA sequence.
 ACCESSION Bf108800
 VERSION Bf108800.1 GI:10938490
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 392)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (InfoImage.lnl.gov) for further information.
 Seq primer: -400P from Glibco
 High quality sequence stop: 358.
 Location/Qualifiers
 1..392
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHSF pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HP8-9W pool 1:
 758280-760583, 772104-774407 Soares NBHPA pool 1:
 304776-306311, 320136-322823, 326280-326653 Soares NBHOF
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 72 a 131 c 153 g 36 t
 ORIGIN
 Alignment_scores:
 Quality: 715.00 Length: 130
 Ratio: 5.500 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.231
 alignment_block:
 US-09-899-513-2 x Bf108800/rev ..
 Align seg 1/1 to reverse of: Bf108800 from: 1 to: 392
 60 ALaasVaAlaAlaLeuAlaMetPheIlAArgSerGIyGlyARgeuI 76
 |||||||
 391 GCCAACGTGGCGGCGCTGCAATGTTACCCGACGCGCGGCGCATGGG 342
 |||||||
 76 yGlnAlaLeuLeuLeuTyrlLeuPheAsnLeuAlaLeuValaSPgluPhep 93
 |||||||
 341 CCAGGCGCTTCTCTACCTGTTCAACCTGGCTGTGTTGAGTAATTTCT 292
 |||||||
 93 heThleuThleuGlnLeuThyrlPheThyTyrleuGlyLeuAlaArg 109
 |||||||
 291 TCACGCTCAGCTGACGTGTGCTCCTACTACTGCGCGCGCCG 242
 |||||||
 110 ArgProAlaTharArgProGlyProProThThThCysProProleAr 126
 |||||||
 241 AGGCGCCCTGCCACGCGCGCGGCGCACTACTACTGTCTCCACTTGGC 192
 |||||||
 126 gArGTrSerSerProArgSerSerAlaCysAlaAlaAlaSerTyrA 143

|||||
 191 GCGGTGCTCTGCGCGCTCATCAGCTGTGCGGCTTGCCTAG 142
 |||||||
 143 laValProGlyProGlyArgLeuProAlaTrpProGlyAlaTyrGlyAla 159
 |||||||
 141 CGGTCCCGGCGCGACGCGCTCCCGCTGCGCGCTGTCTACGGCGCC 92
 |||||||
 160 ProAlaAlaLeuProAlaProSerProGlyTTPArgAlaTrpProLeup 176
 |||||||
 91 CCGCGCGCGCTGCTGCTGCTTGCCTGTGCGCGCGCTGCGCCCTCC 42
 |||||||
 176 oAlaTrpSerThrAlaGlyGlnAlaArgGlyTrpPro 189
 |||||||
 41 TCGCTGAGACCGCTGCGCACTGCGGCGCTGCGCTCCG 2
 |||||||
 seq_name: gb_est1:AM072531
 seq_documentation_block:
 LOCUS AM072531 432 bp mRNA linear EST 20-OCT-2000
 DEFINITION xa08e08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:2567750 3' similar to contains PTR5.b2 TAR1 repetitive
 element.; mRNA sequence.
 ACCESSION AM072531
 VERSION AM072531.1 GI:6027529
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 432)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (InfoImage.lnl.gov) for further information.
 Possible reversed clone: polyt not found
 Seq length: 872 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 432.
 Location/Qualifiers
 1..432
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2567750"
 /clone_id="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CGAP-GCHI) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 84 a 145 c 164 g 38 t 1 others
 ORIGIN
 Alignment_scores:
 Quality: 689.50 Length: 146
 Ratio: 4.856 Gaps: 1
 Percent Similarity: 97.260 Percent Identity: 95.890
 alignment_block:
 US-09-899-513-2 x AM072531/rev ..

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2709445"
/clone_1lb="NCI_CGAP_Sub1"

```

86 LeuAlaIeuValAspIleuPhehSerLeuThrIeuThrIeuGlnIeuTrpLeuTh 102
332 CTGGCTATGAGTGAATGATTATTCATGCTAACGCTGACAGCTGTGGCTCAC 283
102 rTyrTyLeuGlyLeuAlaArgArgProProAlaThrArProGlyProp 119
282 ATACTACTAGGGCTGGCCCCGGGGGGCCCCCTCC .ACGGGGCCCCGGGGCCAC 234
119 roThrCysPropMetArgArgTrpSerSerProArgSerSerAla 135
233 ATAAATAGCGACCACTTATGCGGGCGGTGGTTCGCCGCTCATACAGC 184
136 CysAlaAlaAlaAlaSerTyAlaValProGlyProGlyArGleuProAl 155
183 TGTGCGCGTCCGCTTCGTACCGCGGTCGCCGAGGGCGGCTGCCGC 134
152 aATPProGlyAlaTyGlyAlaProArgAlaIeuProAlaProSerProG 168
133 ATGGCCCGGTGCTACAGGGGCCCGCGCGCGCTGCTCGCTTCGGCTG 84
169 TyTrpArgAlaATrProIeuProAlaTrpSerTrhAlaGlyGlnAlaArg 185
83 GCTGGCGGGCTTGGCCCCCTCCCTGCTGAGACACCGCTGGGCAAGCTCG 34

186 G1YTPProProArgrProSerArgrPro 196
 |||||
 33 GGCTGGCTCCGCCACGGTGGCTTGGCGCCG 1

seq_name: gb_est2:BF307868

seq_documentation_block:

LOCUS BF307868 1083 bp mRNA linear EST 21-NOV-2000
 DEFINITION 601890683F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4131836 5',
 mRNA sequence.

ACCESSION

BF307868

VERSION

BF307868.1

KEYWORDS

EST.

SOURCE

ORGANISM

Human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 1083)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs.femail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L1CM1030 row: e column: 21

High quality sequence stop: 466.

Location/Qualifiers

1..1083

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4131836"

/clone_id="NIH_MGC_17"

/tissue_type="rhabdomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Origin: muscle; Vector: pOTB7; Site_1: EcoRI;

Site_2: XhoI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

218 a 363 c 318 g 184 t

alignment_scores:

Quality: 423.50 Length: 220

Ratio: 2.786 Gaps: 17

Percent Similarity: 69.091 Percent Identity: 57.273

alignment_block:

US-09-899-513-2 x BF307868 ..

Align seg 1/1 to: BF307868 from: 1 to: 1083

2 G1YAlaAlaPheValAlaSerLeuArgSerAnLeuSerSerAlaThrSe 18

|||||

147 GGCAGAGCGCTGGGACACCTCTCTGGAGAGACCTTACCAAGGTCAGAC 196

18 rArgSerGluMetAsnSerSer.ValG1YAspLeuG1YAlaG1YG1YCs 34

|||||

197 G.....ATGACACGACAGTCTCGGGACCTGGGTGTGGCGGCTGC 237

35 SerLeuTPAspAspProAlaArgPheIleValProAlaAlaTyrAl 51

|||||

238 ACCCTCTGGATGACCTCTCTGCTTCATCTGCTGCTCCGCGGCTATGC 287

51 AlenAlaLeu...G1YLeuG1YLeuProAlaAsn.ValAlaAlaLeuAla 66
 |||||

288 CTGTGCACTTGGGCGCTGTGGGCTGCGACGACATCTGCGCGGCTTGCA 337

67 MetPheIleArgSerGly...G1YArgLeuG1YAlaAlaLeuLeuLeu 82

|||||

338 ARGTTATCTGACCGCGGCTGGGCGGCTGGGCGACGCGCTGCTTCTTA 387

82 rLeuPheAsnLeu...AlaLeuValAspGluPhePheThrLeuThr 97

|||||

388 CCTGTTCACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437

97 eugG1YLeuPheThrThrThrThrThrThrThrThrThrThrThrThr 113

|||||

438 CCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487

114 Thr.ArgProGlyProProThrThrThrThrThrThrThrThrThrThr 129

|||||

488 ACTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 537

129 rSerProArgSerSer.AlaCysAlaAlaAlaAlaSerTyrAlaValP 145

|||||

538 TCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGC 587

145 ro.G1YProG1YArgLeuProAlaTyrProG1YAlaTyrG1YAlaProAr 161

|||||

588 CAGGACACAGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 637

161 gAla.....LeuProAlaProSerProGlyTyr...ArgAla.TyrPro 174

|||||

638 TGCTGTACAGTACGACGACGACGACGACGACGACGACGACGACGACG 687

175 LeuProAlaTyrSerThrAlaG1YAlaArgG1YTrpProProProAr 191

|||||

688 CTTCCTCTGCTGTTGAGACACCTGCTGCGGCGGCGGCGGCGGCGG 728

191 gTPP.....ProSerArgProProSerCys 199

|||||

729 CTGGACTTGGGCGGCGGAGAGGCTTGCATGCTGCTGCTGCTGCTGC 771

seq_name: gb_est1:AW014455

seq_documentation_block:

LOCUS AW014455 220 bp mRNA linear EST 10-SEP-1999

DEFINITION UT-H-B10-aat-a-07-0-UI-s1 NCL_CGAP_Sub1 Homo sapiens cDNA clone

IMAGE:2709396 3', mRNA sequence.

ACCESSION AW014455

AW014455

VERSION AW014455.1 GI:5863212

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 220)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs.femail.nih.gov

Oligo-dT track not found. Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: NCI-CGAP clone distribution

information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbp/image/image.html

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1..220

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2709396"

FEATURES

source

```

/clone_lib="NCI_CGAP_Sub1"
/lab_host="DHI0B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI. The
NCI_CGAP_Sub1 library is a subtracted library derived from
BT. BI constitutes a mixture of 21 normalized or
subtracted NCI_CGAP libraries: NCI_CGAP_Co4,
NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16,
NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3,
NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8,
NCI_CGAP_Co11, NCI_CGAP_Le12, NCI_CGAP_Br23, NCI_CGAP_Lu5,
NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_Gc4, NCI_CGAP_Co6,
NCI_CGAP_Br25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with a
driver whose composition is detailed below: NCI_CGAP_Kid3
pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
Clones 1322376-1323911, 1456008-1456775, 1500552-1502855
) NCI_CGAP_Kid5 pool 1 LAM 3338-3342, 3722-3725,
3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LAM
3575-3582, 3851-3854 (IMAGE Clones 1414920-1417911,
1520904-1522439) NCI_CGAP_Gc4 pool 1 LAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1
LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 121928-1220615)
NCI_CGAP_Co10 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351) The resulting
subtracted library contained 530,000 recombinants.
Subtraction was performed as previously described [Bonaldo
& Lennon & Soares (1996)]. Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG LIB=NCI_CGAP_Le12
TAG_TISSUE=leiomysarcoma
TAG_SEQ=AAATC"

```

BASE COUNT 34 a 81 c 84 g 21 t

ORIGIN

alignment_scores:

Quality: 351.00 Length: 73

Ratio: 5.400 Gaps: 0

Percent Similarity: 89.041 Percent Identity: 86.301

alignment_block:

US-09-899-513-2 x AM014455/rev ..

align seg 1/1 to reverse of: AM014455 from: 1 to: 220

```

124 PROMELARGATGTPSerSerProArgSerSerAlaCysAlaAlaAlaAl 140
|||||
219 CCTATGGGGGGGTGTCGCGCGATACAGAGGTGTCGCGCGCGGC 170
|||||
140 aSerTtAlaAlaValProGlyProGlyArgLeuProAlaTrpProGlyAla 157
|||||
169 TTCTACGCGGTCCGGGCGCCAGGCGGATGTCGCGCGCGCGTGC 120
|||||
157 yrgYAlaProArgAlaLeuProAlaProSerProGlyTtPArgAlaTrp 173
|||||
119 ACGGCGCCCGCGAGCTGCGTGCATTTGCGCTGCGGCGGCGCTGG 70
|||||
174 ProLeuProAlaTrpSerThrAlaGlyAlaAlaArgGlyTtPProPr 190
|||||
69 CCCCTCCCTGCTGAGACGATGCGCAACCTCGGGCGTACATCCGCC 20
|||||
190 CATGTTPProSerArgPro 196
|||||
19 ACGGTGCGCGTCCGCGCGCG 1
|||||
seq_name: gb_est1:AM138178

```

```

seq_documentation_block:
LOCUS AM138178 189 bp mRNA linear EST 29-OCT-1999
DEFINITION UT-H-BI-acv-e-11-0-UT.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2716124 3', mRNA sequence.
ACCESSION AM138178
VERSION AM138178.1 GI:6142578
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 189)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/ILUML at:
www-bio.llnl.gov/dbp/image/image.html
Seq primer: M13 forward
POLYA-No.

FEATURES
Source
1..189
location/Qualifiers
1..189
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2716124"
/clone_lib="NCI_CGAP_Sub3"
/lab_host="DHI0B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI. The
NCI_CGAP_Sub3 library is a subtracted library derived from
the NCI_CGAP_Sub1 library, which is a subtracted library
derived from BT. BI constitutes a mixture of 21
normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4,
NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_Co11, NCI_CGAP_Le12,
NCI_CGAP_Br23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_Gc4, NCI_CGAP_Co6,
NCI_CGAP_Br25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911, 1456008-1456775
, 1500552-1502855) NCI_CGAP_Kid5 pool 1 LAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1
LAM 3575-3582, 3851-3854 (IMAGE Clones 1414920-1417911,
1520904-1522439) NCI_CGAP_Gc4 pool 1 LAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1
LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 121928-1220615)
NCI_CGAP_Co10 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described [Bonaldo, Lennon &
Soares (1996)]. Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG LIB=NCI_CGAP_Gc4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"

```

BASE COUNT 28 a 66 c 69 g 26 t

ORIGIN

alignment_scores:

Quality: 163.50 Length: 212
 Ratio: 1.858 Gaps: 12
 Percent Similarity: 41.509 Percent Identity: 32.075

alignment_block:
 US-09-899-513-2 x BF312953/rev ..

Align seg 1/1 to reverse of: BF312953 from: 1 to: 843

```

26 ValGlyAspLeuGly.....ValGlyGlyCys..... 34
|||||
588 GTAGAGCCGATAGTGGCTTGGCGCTTGGCGAGGCTGCTCC 539
|||||
35 .....SerLeuTrpAspAspProAlaArgPheIleV 45
|||||
538 TACCGGAACAGCGCTCGGCGCATCTGGCCAGACCTCCAGAGGACAG 489
|||||
45 AlValProAlaAlaTyrAlaLeuAlaLeuGlyLeuGlyLeuProAlaAsn 61
::: ||||| ::|||
488 CAGAGCCAGGCGCACAGGCT...GCCCTGGGGGCCAGCCCTGCACCGGCT 442
|||||
62 ValAlaAlaLeuAlaMetPheIleArgSerGlyGlyArgLeuGlyAla 78
|||||
441 GTCTCAGCCCGCCGCCGTAGCTCTGCACCTGTGGCCACTTCTTCACAGC 393
|||||
78 AlLeuLeuLeuTyrLeuPheAsnLeuAlaLeuValAspLupPhePheHrL 95
|||||
392 T..... 392
|||||
95 eutHrLeuGlnLeuTyrPheThr.....TyrTyr 104
|||||
391 .....CTGCCCAATGCGCGACAGGCTCTCCAGCTCGTGTCTGTGTC 347
|||||
105 LeuGlyLeuAlaArgArgProProAlaThrArgPro..... 116
|||||
346 TGGCGTTTGGCTGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 297
|||||
117 .....GlyProProThrThrCysProProMetArgArgTyrPS 129
|||||
296 ACCGTGAGGAGGCGCGCGCTCAAGGCTGTCTCTCGGAGGAGATG.... 251
|||||
129 erSerProArgSerSerAlaCysAlaAlaAlaSer.....Tyr 142
|||||
250 .....GCCTCCAGCTCTGCGGAGCTGCGGCGCCAGGCGCCCTGAGAC 209
|||||
143 AlValAlaProGlyProGlyArgLeuProAlaTrp.....ProGlyAlaTyr 157
|||||
208 CGTAGGCGCGGCGCGGAGGAGGAGCTTCATGAGACTGTCGCGGTGTTC 159
|||||
157 rGlyAlaProAlaArgAlaLeuProAla.....ProSerProGlyTyrPA 171
|||||
158 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 109
|||||
171 rGlyAlaTrpPro.....LeuProAlaTrpSerThrAlaGlyAlaArg 185
|||||
108 CAGCTGATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 77
|||||
186 GlyTrpProProAlaArgTrpProSerArgProPro 197
|||||
76 .....CTCTCCACATCTTGGAGAGCGCTCTCTCT 47
|||||
seq_name: gb_est2:BE259788
seq_documentation_block:
LOCUS BE259788 575 bp mRNA linear EST 26-OCT-2000
DEFINITION 601152574F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3508555 5',
mRNA sequence.
ACCESSION BE259788
VERSION BE259788.1 GI:9130544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
 1 (bases 1 to 575)
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 Plate: LICM190 row: c column: 20
 High quality sequence stop: 571.

FEATURES
 source
 1..575
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3508555"
 /clone_id="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Library constructed by Ling Hong
 in the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using Zap-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

BASE COUNT 128 a 155 c 231 g 61 t

alignment_scores:
 Quality: 161.00 Length: 172
 Ratio: 1.872 Gaps: 9
 Percent Similarity: 50.000 Percent Identity: 36.047

alignment_block:
 US-09-899-513-2 x BE259788/rev ..

Align seg 1/1 to reverse of: BE259788 from: 1 to: 575

```

40 ProAlaArgPheIleValValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
|||||
567 CCGGCGAGCGGCTGTCTCAGCCCGCGGTAGTGCCTGCGAGCTGGGGGCC 518
|||||
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGly 72
|||||
517 AGCCCTGCCAGCGCGGTCTCAGCCCGCGGTAGTGCCTGCGAGCTGGTG 468
|||||
73 GlyArgLeuGlyAlaLeuLeuLeuTyrLeuPheAsn.LeuAlaLeuV 89
|||||
467 GGCACATTTCTCCACACTCTGCTCAATGAGCGGACAGCGCTCCACCTCC 418
|||||
89 AlAspLupPhePheHrLeuThrLeuGlnLeuTyrPheThrTyrTyr 105
|||||
417 TGGTCTCTGCTGCTGGGTTGGCGCTCTCTCTCTCTCTCTCTCTCTCT 368
|||||
106 GlyLeuAlaArgArgProProAlaThrArgProGlyProProThrHrty 122
|||||
367 GGGGTAGCTTCAGAGACTGGAGGGCA.....GGCCCGCTCAAGGCTG 324
|||||
122 sProProMetArgArgTyrSerSerProArgSerSerAlaAlaAla 139
|||||
323 TCCTCCCGGAGAGATG.....GCCGCCAGCTCGTCGACAGCTG 286
|||||
139 lAlaSer.....TyrAlaValProGlyProGlyArgLeuProAla 152
|||||
285 CGGCCAGGCGCCCTGAGACCGGTAGCCCGGCGCGGAGGAGAGAGCTTCA 236
|||||

```

